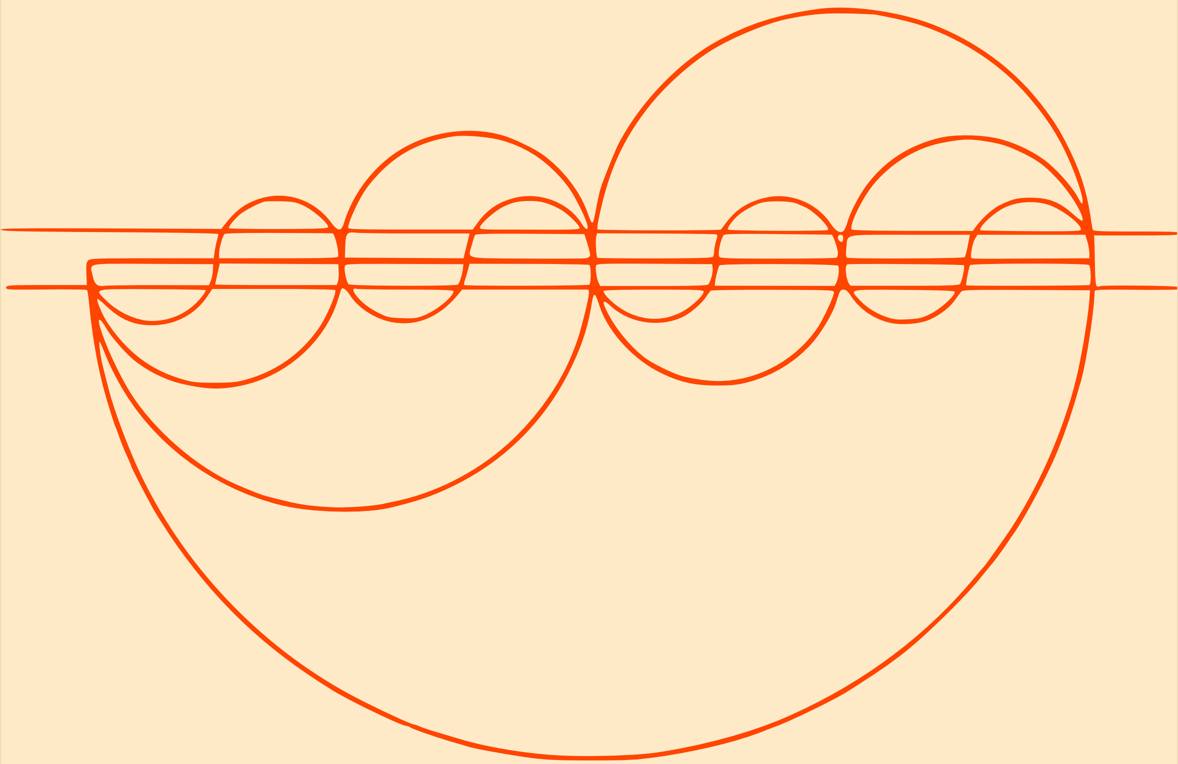
TOPICS IN MATHEMATICS

*Ya. S. Dubnov*

# Mistakes in Geometric Proofs



TOPICS IN MATHEMATICS

# Mistakes in Geometric Proofs

## Ya. S. Dubnov

*Translated and adapted from the second Russian edition (1955) by*

ALFRED K. HENN *and* OLGA A. TITELBAUM

SURVEY OF

RECENT EAST EUROPEAN MATHEMATICAL LITERATURE

*A project conducted by*

ALFRED L. PUTNAM *and* IZAAK WIRSZUP

*Department of Mathematics,*
*The University of Chicago, under a*
*grant from the National Science Foundation*

D. C. HEATH AND COMPANY BOSTON

Library of Congress Catalog Card Number: 61-17840

Copyright © 1963 by The University of Chicago
No part of the material covered by this copyright may be reproduced in any form without written permission of the publisher.
Printed in the United States of America
(6 L 2)

## PREFACE TO THE AMERICAN EDITION

THIS BOOKLET presents examples of faulty geometric proofs, some of which illustrate mistakes in reasoning that a student might make, while others are classic sophisms. Chapters 1 and 3 present these faulty proofs, and then Chapters 2 and 4 give detailed analyses of the mistakes.

Naturally, in order to read this booklet, the reader must be acquainted with plane geometry. Only an acquaintance with theorems concerning parallel and perpendicular lines and polygons is needed for Chapters 1 and 2. Chapters 3 and 4 contain more advanced material and presuppose some knowledge of the simpler properties of circles, the concept of limit, trigonometry, and some solid geometry.

It is suggested that the reader first examine the examples of incorrect proofs given in Chapters 1 and 3. He should attempt to discover the mistakes in these examples by himself before reading Chapters 2 and 4. Portions of the text appearing in fine print, as well as many of the footnotes, may be omitted on first reading; these are intended primarily for the more advanced reader.

CONTENTS

# Introduction

Forty years ago in an article dealing with the teaching of geometry, a then well-known Russian mathematician and teacher, N. A. Izvolskii, reported a conversation which he had had with a school-girl acquaintance who had just been studying geometry for one year. The teacher asked her what she remembered from the geometry course. After thinking for a long time, the girl was unfortunately unable to remember anything. The question was then recast in another form: "What did you do during the year in your geometry lessons?" This brought the very quick reply, "We did proofs."

This answer reflects an idea widely held among pupils—in arithmetic problems are solved, in algebra equations are solved and formulas are derived, but in geometry theorems are proved. This conception of mathematics was correct a long time ago, but in present-day mathematical studies, theorems followed by proofs are encountered equally often whether one is dealing with numbers or diagrams. Moreover, problems are solved in all fields of mathematics, and in geometry we often solve equations. It was different 2,000 years ago when Euclidean geometry, which still forms the basis of the high school geometry course, was being created. From that time up to the present day, geometry (but not the other branches of mathematics) has been expounded in textbooks in the form of a chain of theorems (some of which are called lemmas or corollaries). These are constructed according to a plan which is so well known that we limit ourselves to a short reminder. Each theorem contains a condition ("Given . . .") and a conclusion ("Prove . . ."); in the proof we may rely only on axioms or on theorems already proved.

The part which drawings play in proofs is well known; they clarify not only the content of the theorem but also the course of the proof. Sometimes several drawings are required for a single theorem, because the proof may vary with the relative positions of the parts of a figure. For example, the proof of a theorem about an angle inscribed in a circle usually involves three possibilities: the center of the circle lies on a side of the angle, inside of the angle,

or outside of it. It is important to exhaust all possible arrangements of the parts of the figure; the omission of any one arrangement to which the exhibited reasoning cannot be applied, of course, invalidates the entire proof—it may be precisely for that one arrangement that the theorem turns out to be false.

The role of the drawing should be neither exaggerated nor underestimated. To regard it as an indispensable part of the proof would be an exaggeration. Theoretically speaking, a geometric proof does not need drawings; not using them even has the advantage of removing any reliance on what is "self-evident" from the drawing, which sometimes is only apparently "self-evident" and is a source of error. In practice, however, dispensing with the drawing might lead to the same difficulties as we would experience if, for instance, we were to try to perform computations with numbers having several digits completely in our heads, or if we were to play chess without looking at the chess board; the danger of making a mistake would be increased considerably.

In speaking of the help afforded by a drawing in developing a proof, I have in mind, of course, a good drawing carefully executed. A bad drawing can be a hindrance. In this booklet the reader will encounter not only correct drawings, but also others which are somewhat distorted. This is done deliberately; we are concerned here with faulty proofs, and these sometimes result from inaccurate drawings.

In Chapters 1 and 3, a number of examples of faulty geometric proofs will be given. These "proofs" will then be analyzed in Chapters 2 and 4.

Among these propositions there are some whose falsity will be immediately apparent to the reader, for instance, "A right angle is equal to an obtuse angle." Here our task is to discover the mistake in the proof. Such proofs of deliberately incorrect assertions have been known from ancient times as "sophisms."

In other examples the reader will not know in advance whether the assertion proved is true or false if he has not come across it before. Here our task is more complicated; we must discover that the proof is unsound and whether or not the assertion is erroneous. To discover only the first is not sufficient; it is in fact possible to base a correct assertion on faulty arguments. For instance, from the incorrect equation $3 + 5 = 12$ it is possible to deduce correctly that $3 + 5$ is an even number.

Finally, examples will be given of proofs whose invalidity stems from the fact that the proposition asserted has nothing to do with the given data. How this may come about I shall attempt to explain by an example which is remote from geometry and science in general.

The following facetious problem is well-known: "A steamer is situated at latitude 42°15′ N. and longitude 17°32′ W. [The figures are taken at random; usually further data is added which complicates the conditions.] How old is the captain?" For our purpose let us alter the question of the problem somewhat. "Is the assertion correct that the captain is more than 45 years old?" It is clear to everyone that it is impossible to draw such a conclusion from the data given in the conditions of the problem, and that any attempt to prove the assertion concerning the age of the captain is destined to end in failure. Moreover, it is possible to *prove* that it is impossible to prove this assertion. Actually the steamship company, about which we learn nothing from the data of the problem, may chart a course which passes through the geographic point indicated and assign to the voyage a captain of this or that age, assuming that the company has captains of any age available for such trips.

In other words, it is possible to assume that the captain is younger than 45 years old without in any way contradicting the data concerning the latitude and longitude. It is another matter if the conditions of the problem contain other data as well, such as the name of the ship and the date on which it passes through the point indicated; one might then hope to establish the identity of the captain and then his age by using the ship's log. Thus, there exist assertions whose validity may or may not be proved, depending on the data which we have at our disposal for obtaining the proof.

Returning more nearly to our subject, let us ask, "Is it true that the sum of the angles of any triangle is equal to two right angles?" Every schoolboy who has studied the chapter on parallel straight lines in a geometry textbook is acquainted with the proof of this important theorem, but few know its history, which goes back 2,000 years. The proof is based on the properties of angles formed by a line intersecting parallel straight lines, and these properties are based in turn on the so-called "parallel postulate": *Only* one straight line can be drawn parallel to a given line through a given

point not on this line.[1] Since the time of Euclid, efforts have been made to turn this axiom into a theorem, that is, to prove it only on the basis of the other axioms and on assertions which both in Euclid and in our textbooks do not depend on the parallel postulate. But we cannot "prove" this axiom like a theorem; all such attempts have been unsuccessful and it has been found only that the parallel postulate may be replaced by *equivalent* ones in many different ways. In particular, if we take as an axiom one of the properties of the angles formed by a pair of parallel straight lines and a third line intersecting them, or the theorem about the sum of the angles of a triangle, then the parallel postulate becomes a theorem.

It was not until the eighteen-twenties that the Russian mathematician Nikolai Lobachevskii (1792–1856) discovered the cause for the failure of all attempts to prove the parallel postulate. He constructed an extensive and profound theory of geometry, of which I shall not attempt to give here even the remotest idea. Among other things, this theory shows that it is impossible to prove the parallel postulate, which had been attempted by many scholars up to the time of Lobachevskii (and during his lifetime).

However complex the theory of Lobachevskii, and on the other hand, however naive the problem about the age of the captain, the "proof of the impossibility of the proof" is of the same nature in both problems. In these problems we are given certain data, and we desire to show that a certain conclusion cannot be logically deduced from the data. To do this we find concrete examples (called "models") in which all the given conditions are satisfied, but in some of which the conclusion in question is true and in others the conclusion is false. Applied to the parallel postulate this means that neither the truth nor the falsity of the parallel postulate follows from the other axioms of Euclidean geometry. We now know that any proof for the parallel postulate, or for any other postulate equivalent to it, must be invalid if it is based only on the propositions based on the other axioms. In Chapter 1 we shall give several simple examples of such faulty proofs.

[1] Note that the axiomatic nature of this proposition is based on the word "only." The fact that it is always possible to draw one parallel line can be proved earlier, if only on the basis of the theorem that two perpendiculars to the same straight line are parallel to each other.

# 1. Mistakes in Reasoning within the Grasp of the Beginner

We shall now proceed to give examples of faulty proofs, bearing in mind that their critical analysis will be postponed until Chapter 2. The reader has been forewarned that some of the drawings in this booklet contain distortions which are sometimes not at once apparent.

EXAMPLE 1. *The square whose side is* 21 *cm. has the same area as a rectangle whose sides are* 34 *cm. and* 13 *cm.*

The square $Q$ is divided into two rectangles whose dimensions are $13 \times 21$ and $8 \times 21$ (Fig. 1; "cm." will subsequently be

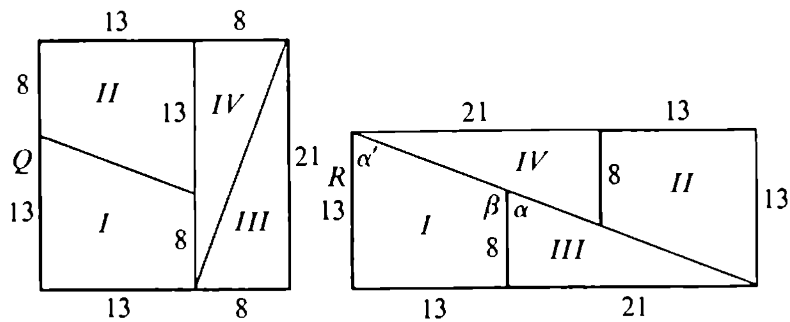



Fig. 1

omitted). The first rectangle is divided into two identical rectangular trapezoids the lengths of whose parallel sides are 13 and 8, and the second rectangle into two congruent right triangles whose legs are 8 and 21. These four parts are rearranged to form the rectangle $R$ shown in Fig. 1 on the right. Corresponding parts of the square and of the rectangle are designated by the same Roman numerals.

Precisely speaking, we place the right triangle *III* next to the rectangular trapezoid *I* in such a way that the right angles at the common side of length 8 are adjacent; a right triangle is formed with legs 13 and $13 + 21 = 34$. An identical triangle is obtained from parts *II* and *IV*, and these two congruent right triangles are put together to form the rectangle $R$ with sides 13 and 34.

The area of this rectangle is equal to

$$34 \times 13 = 442,$$

while the area of the square $Q$, which is made up of the same parts, is

$$21 \times 21 = 441.$$

Where does the extra square centimeter come from? We suggest that the reader carry out an experiment. Cut the square $Q$ out of paper (preferably paper ruled in squares), taking the side of one square to represent 1 cm., dissect $Q$ into four parts, carefully observing the dimensions indicated, and rearrange these parts to give the rectangle $R$.

EXAMPLE 2. *A proof of the parallel postulate.*

Given a straight line $AB$ and a point $C$ outside it, prove that through the point $C$ only one straight line parallel to $AB$ can be drawn. Using a familiar construction, drop a perpendicular from the point $C$ to the straight line $AB$ (Fig. 2; in this figure and

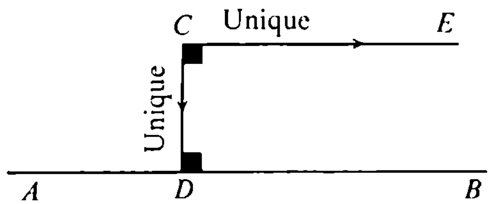


Fig. 2

in many subsequent ones, right angles will be marked by solid squares). To this perpendicular erect a perpendicular $CE$ from the point $C$. This second perpendicular will be parallel to the straight line $AB$ by virtue of the theorem that two perpendiculars to the same straight line are parallel. Note that it is legitimate to refer to this theorem here, because it can be proved without using the parallel postulate. But it is possible to drop only *one* perpendicular from a given point to a given straight line, and it is possible to erect only *one* perpendicular to a straight line from a point lying on it; both these facts can be proved without using the parallel postulate. Therefore, the parallel line $CE$ obtained is *unique.*

Example 3. *If two parallel straight lines are intersected by a third line, the sum of the interior angles lying on the same side of the third line is equal to* 180° (*proof not based on the parallel postulate*).

Say $AB \parallel CD$, and let the line $EF$ intersect these two lines (Fig. 3). The interior angles are designated by numbers in the drawing.

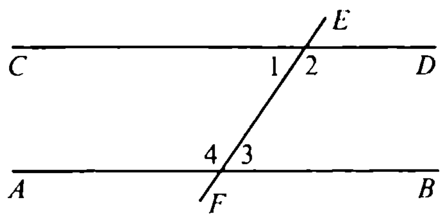



Fig. 3

Three cases are possible:[1]

1. The sum of the interior angles lying on the same side of the line $EF$ is $> 180°$.
2. The sum of the interior angles on the same side of the line $EF$ is $< 180°$.
3. The sum of the interior angles on the same side of the line $EF$ is $= 180°$.

In the first case we have

$$\angle 1 + \angle 4 > 180°, \quad \angle 2 + \angle 3 > 180°;$$

therefore,

$$\angle 1 + \angle 2 + \angle 3 + \angle 4 > 360°.$$

But the sum of the four interior angles is equal to two straight angles, that is, 360°. This contradiction shows that the first assumption must be discarded. For the same reason we must also abandon the second assumption, as it would lead to the conclusion that the sum of the four interior angles is less than 360°. The third assumption is the only possible one (it does not lead to a contradiction); this proves the theorem.

[1] Here and subsequently, when talking about possible assumptions or possible cases, we do not by any means assert that they are all actually possible under the conditions of the given example. On the contrary, time and again it will happen that what is at first assumed to be a possible case later turns out to be spurious—contrary to the conditions or to what is taken as established; this often happens in indirect proofs. Thus, we are talking throughout about so-called "*a priori* possibilities," that is, about possibilities which present themselves beforehand, prior to taking into account the other conditions of the problem.

EXAMPLE 4. *The sum of the angles of a triangle is equal to* 180° (*proof not based on the parallel postulate*).

Divide the arbitrary triangle $ABC$ into two triangles by means of a line segment drawn from the vertex, and denote the angles by numbers as in Fig. 4. Let $x$ be the sum of the angles of a triangle, unknown as yet; then

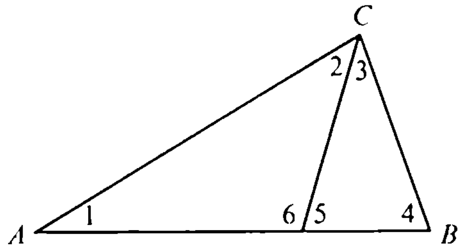


Fig. 4

$$\angle 1 + \angle 2 + \angle 6 = x,$$
$$\angle 3 + \angle 4 + \angle 5 = x.$$

Adding these two equalities, we obtain

$$\angle 1 + \angle 2 + \angle 3 + \angle 4 + \angle 5 + \angle 6 = 2x.$$

But the sum $\angle 1 + \angle 2 + \angle 3 + \angle 4$ is the sum of the angles of the triangle $ABC$, that is, it is also $x$; and the angles 5 and 6, being adjacent angles, have a sum equal to 180°. Thus, for finding $x$ we have the equation $x + 180° = 2x$, from which it follows that $x = 180°$.

EXAMPLE 5. *There exists a triangle the sum of whose angles is equal to* 180° (*proof not based on the parallel postulate*).

We shall begin with a historical note. In the eighteenth and at the beginning of the nineteenth century some mathematicians attempted to show that it is possible to speak about the sum of the angles of a triangle without referring to the parallel postulate. It was established that the sum of the angles of a triangle cannot be greater than 180°. There remained three possibilities: (1) for all triangles this sum is equal to 180°, (2) for all triangles it is less than 180°, (3) it is sometimes equal to and sometimes less than 180°. It was subsequently found that the third of these possibilities can be excluded. Efforts were then concentrated on obtaining at least one example of a triangle the sum of whose angles is equal to 180°. We shall now describe one attempt to achieve this; if it should succeed, the parallel postulate would become superfluous.

As the sum of the angles of a triangle does not exceed 180°, let the triangle $ABC$ (see Fig. 4) be a triangle the sum of whose angles is greatest; we shall designate this sum by $\alpha$. If there are several such triangles, we shall take one of them at random. Thus, the sum

of the angles of any other triangle will not exceed $\alpha$; therefore, retaining the notation of Fig. 4, we have

$$\angle 1 + \angle 2 + \angle 6 \leq \alpha, \quad \angle 3 + \angle 4 + \angle 5 \leq \alpha.$$

From this we obtain $\angle 1 + \angle 2 + \angle 3 + \angle 4 + \angle 5 + \angle 6 \leq 2\alpha$; but by assumption $\angle 1 + \angle 2 + \angle 3 + \angle 4 = \alpha$, and moreover, $\angle 5 + \angle 6 = 180°$; consequently, $\alpha + 180° \leq 2\alpha$, $\alpha \geq 180°$. And inasmuch as $\alpha$ cannot be greater than 180°, we must have $\alpha = 180°$ this is, the sum of the angles of triangle $ABC$ is 180°.

EXAMPLE 6. *All triangles are isosceles.*

Let $ABC$ be an arbitrary triangle (Fig. 5, 6, or 7); construct the bisector of the angle $C$ and the perpendicular bisector of the side $AB$. We shall consider the different relative positions of these lines.

Case 1. *The bisector of C and the perpendicular bisector of AB do not intersect*; they are either parallel or they coincide. The bisector of $\angle C$ will then be perpendicular to $AB$; that is, it will coincide with the altitude. Then triangle $ABC$ is isosceles ($CA = CB$).

Case 2. *The bisector of $\angle C$ and the perpendicular bisector of AB intersect inside the triangle ABC*, say at the point $N$ (Fig. 5). Since

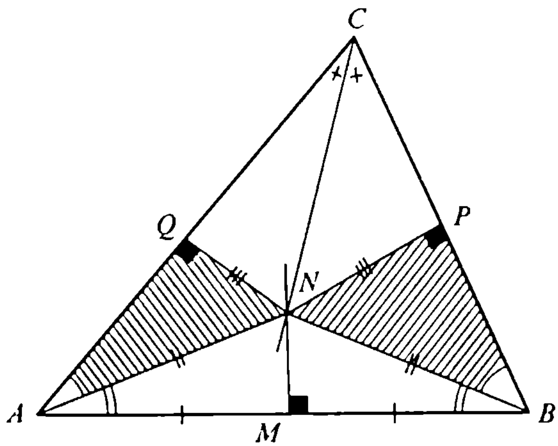


Fig. 5

this point is equidistant from the sides of the angle $ACB$, the perpendiculars $NP$ and $NQ$ to $CB$ and $CA$, respectively, will be equal. But the point $N$ is also equidistant from the end points of the line segment $AB$; that is, $NB = NA$. Then the right triangles $NPB$ and $NQA$ are congruent (leg-hypotenuse); hence, $\angle NAQ = \angle NBP$.

Adding to these equal angles the angles $NAB$ and $NBA$, which are equal to each other since they are base angles of the isosceles triangle $ANB$, we obtain $\angle CAB = \angle CBA$; therefore, the triangle $ABC$ is isosceles.

Case 3. *The bisector of $\angle C$ and the perpendicular bisector of $AB$ intersect on $AB$*, that is, at the mid-point $M$ of $AB$. This means that the median and the angle bisector from vertex $C$ coincide; it follows that the triangle is isosceles.

*Note.* The reader is warned against a possible mistake. It is well known that the median and the angle bisector from the vertex opposite the base of an isosceles triangle coincide. We are referring here not to this, but to the converse assertion: "If the median and the angle bisector from the same vertex of a triangle coincide, the triangle is isosceles." This converse theorem is also true, but the reader may find it difficult to prove; we shall, therefore, indicate one possible method. In the triangle $ABC$ suppose $CM$ is both the median and the angle bisector. Dropping the perpendiculars $MP$ and $MQ$ to the sides $CB$ and $CA$ from the point $M$, we obtain the congruent right triangles $MPB$ and $MQA$. Fig. 5 may be used, taking the points $M$ and $N$ to coincide; the straight line $MN$ then becomes superfluous. Then from the equality of the angles $MBP$ and $MAQ$ we conclude that the triangle $ABC$ is isosceles. This reasoning will be incomplete if we do not show that the points $P$ and $Q$ fall on the sides $CB$ and $CA$ themselves, and not on their extensions. One of these points might fall on the corresponding extension if either the angle $A$ or the angle $B$ were obtuse. Say, for instance, the angle $B$ is obtuse so that the point $P$ lies on the extension of $CB$. As before, we obtain $\angle MAQ = \angle MBP$; but this now leads to a contradiction, as the first of these angles is interior to the triangle $ABC$, while the second is exterior and not adjacent to the first. (Why does this lead to a contradiction?)

Case 4a. *The bisector of $\angle C$ and the perpendicular bisector of $AB$ intersect outside the triangle $ABC$; the perpendiculars dropped*

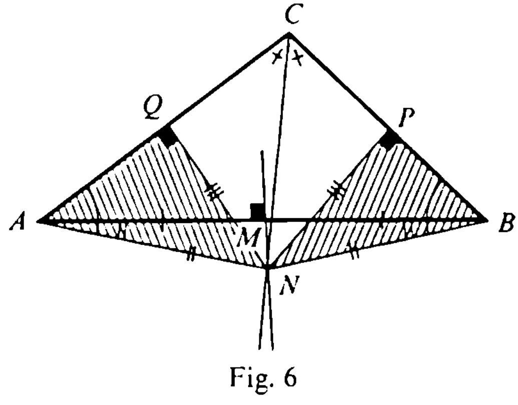



Fig. 6

*from their point of intersection N to the sides CB and CA fall on these sides and not on their extensions* (Fig. 6). As before, we obtain the congruent triangles $NPB$ and $NQA$ and the isosceles triangle $ANB$. The angles at the base $AB$ of the triangle $ABC$ are now equal, being the difference (not the sum, as in case (2)) of corresponding equal angles.

Case 4b. *The bisector of $\angle C$ and the perpendicular bisector of $AB$ intersect outside the triangle $ABC$; the perpendiculars dropped from their point of intersection $N$ to the sides $CB$ and $CA$ fall on their extensions* (Fig. 7). The same constructions and reasoning lead to the conclusion that the exterior angles at vertices $A$ and $B$ of the triangle $ABC$ are equal. From this it follows immediately that the interior angles at $A$ and $B$ are equal; consequently, $CA = CB$.

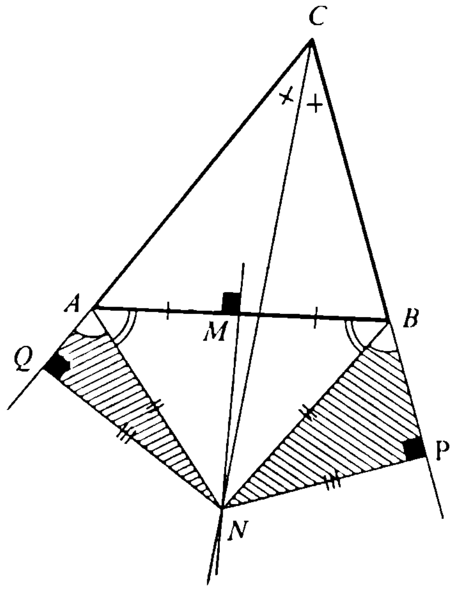


Fig. 7

Example 7. *A right angle is equal to an obtuse angle.*

From the end points of the line segment $AB$ (Fig. 8 or 9) draw two equal line segments $AC$ and $BD$, lying on the same side of $AB$ and forming with it the right angle $DBA$ and the obtuse angle $CAB$; we shall now prove that these two angles are equal. By joining $C$ and $D$ we obtain the quadrilateral $ABDC$, whose sides $AC$ and $BD$ are clearly not parallel; the same applies to the sides $AB$ and $CD$, for otherwise $ABDC$ would be an isosceles trapezoid with unequal angles at the base $AB$. Construct the perpendicular bisectors of each of the line segments $AB$ and $CD$. Since the line segments $AB$ and $CD$ are not parallel, their perpendicular bisectors will not be parallel either and will not coincide, but will intersect at some point $N$.

Let us examine the possible cases.

Case 1. *The point N lies "above" the straight line AB*, strictly speaking, on the same side of $AB$ as the quadrilateral $ABDC$ (see Fig. 8, where the point $N$ is situated in the interior of the quadri-

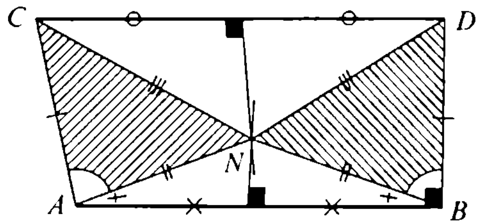



Fig. 8

lateral). Join this point with all the vertices of the quadrilateral. Since it is equidistant from the end points of the line segment $AB$ and likewise from the end points of $CD$, the triangles $NAC$ and $NBD$ are congruent because the three pairs of corresponding sides are equal. From this it follows that $\angle NAC = \angle NBD$. Adding the angle $NAB$ to the first angle and the angle $NBA$ to the second and taking into account that $\angle NAB = \angle NBA$ by the properties of isosceles triangles, we arrive at the equality $\angle CAB = \angle DBA$.

Case 2. *The point N lies on AB*; that is, it is the mid-point of the line segment $AB$. The preceding proof can be simplified; the equality $\angle CAB = \angle DBA$ follows at once from the congruence of the triangles $NAC$ and $NBD$.

Case 3. *The point N lies "below" AB*; that is, it does not lie on the same side of the line $AB$ as the quadrilateral $ABDC$ (Fig. 9).

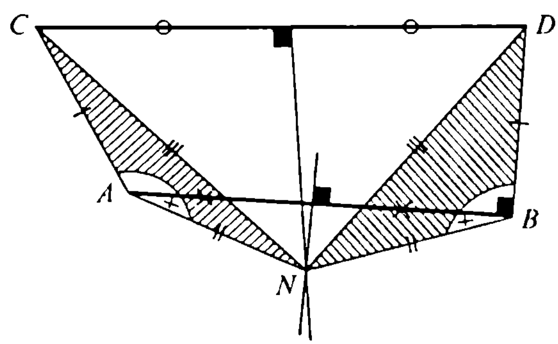



Fig. 9

From the congruence of the triangles $NAC$ and $NBD$ we again obtain $\angle NAC = \angle NBD$, and this time we *subtract* from these angles the angles $NAB$ and $NBA$, which are equal to each other, again obtaining $\angle CAB = \angle DBA$.

EXAMPLE 8. *If two sides and the angle opposite one of them in one triangle are equal to the corresponding parts of another triangle, then these triangles are congruent.*

In the triangles $ABC$ and $A_1B_1C_1$ (Fig. 10, 11, or 12) suppose we are given that

$$AB = A_1B_1, AC = A_1C_1, \text{ and } \angle C = \angle C_1;$$

we shall prove these triangles to be congruent. For this purpose we make use of a method widely used for the proof of the congruence of triangles with equal corresponding sides. Place triangle $A_1B_1C_1$ beside triangle $ABC$ in such a way that the corresponding end points of the equal sides $AB$ and $A_1B_1$ opposite the angles assumed equal coincide. The triangle $A_1B_1C_1$ will then occupy the position $ABC_2$. Joining the points $C$ and $C_2$, let us examine three possible cases.

Case 1. *The straight line $CC_2$ intersects the side $AB$ at a point lying between $A$ and $B$* (Fig. 10). The triangle $ACC_2$ is isosceles;

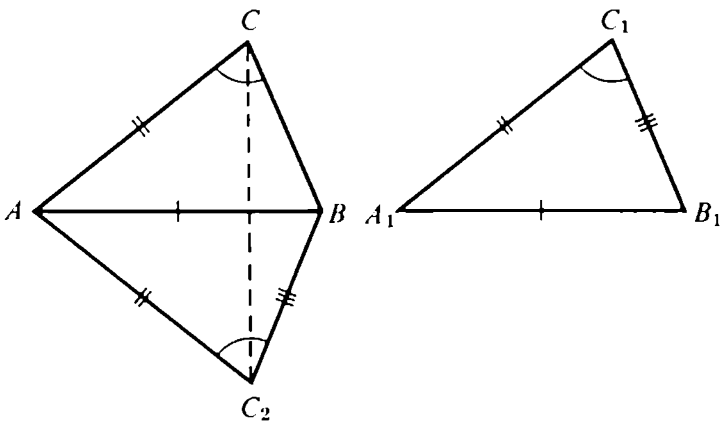



Fig. 10

consequently, $\angle ACC_2 = \angle AC_2C$. If we subtract these equal angles from the angles $ACB$ and $AC_2B$, respectively, which are equal by hypothesis, we obtain

$$\angle BCC_2 = \angle BC_2C.$$

This last equality shows that triangle $CBC_2$ is also isosceles with $CB = C_2B$; therefore,

$$CB = C_2B = C_1B_1,$$

and the triangles $ABC$ and $A_1B_1C_1$ are congruent (side-side-side).

Case 2. *The straight line $CC_2$ intersects the extension of $AB$ beyond $B$* (Fig. 11). The reasoning is the same as before, except that

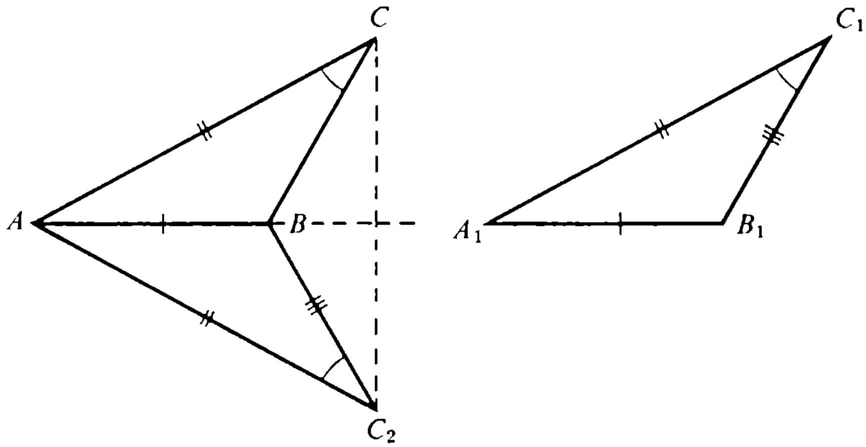



Fig. 11

this time the subtraction is performed in a different order; we subtract the equal angles $ACB$ and $AC_2B$ from the equal angles $ACC_2$ and $AC_2C$.

Case 3. *The straight line $CC_2$ intersects the extension of the side $BA$ beyond $A$* (Fig. 12). The reasoning is the same as in case 1, but

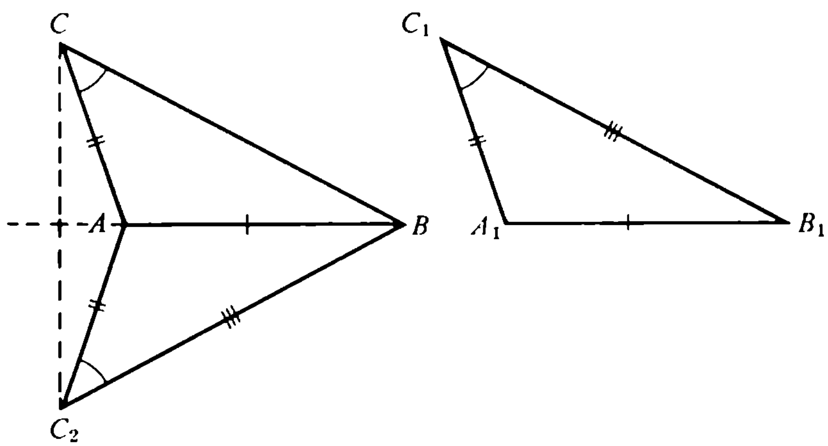



Fig. 12

addition takes the place of subtraction; to the equal angles $ACC_2$ and $AC_2C$ we add the equal angles $ACB$ and $AC_2B$.

EXAMPLE 9. *A rectangle which is inscribed in a square is also a square.*[1]

More precisely, if the rectangle $MNPQ$ (Fig. 13) is inscribed in the square $ABCD$ in such a way that one of the vertices of the rectangle lies on each of the sides of the square, the rectangle will also be a square. In the figure, $M$ lies on $AB$, $N$ on $BC$, $P$ on $CD$, and $Q$ on $DA$.

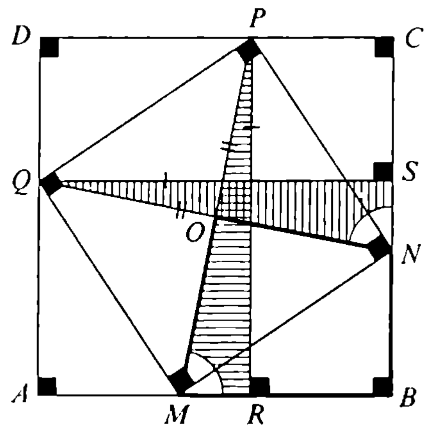


Fig. 13

To prove this, drop perpendiculars $PR$ and $QS$ from $P$ and $Q$ to $AB$ and $BC$, respectively. These perpendiculars are equal to each other, for each is equal to the side of the square $ABCD$. They are the legs of the triangles $PRM$ and $QSN$, whose hypotenuses, as the diagonals of the rectangle $MNPQ$, are also equal to each other. From this it follows that the triangles shaded in the figure are congruent, and hence that

$$\angle PMR = \angle QNS.$$

We now examine the quadrilateral $MBNO$ drawn in heavy lines in the figure, where $O$ is the intersection point of the diagonals of the rectangle $MNPQ$. Its exterior angle at the vertex $N$ is equal to the interior angle at the vertex $M$, so that the two interior angles at the vertices $M$ and $N$ are supplementary. The interior angles at the vertices $B$ and $O$ must also be supplementary, but one of them ($\angle B$) is a right angle, and, consequently, the other ($\angle O$) is also a right angle. Hence, the diagonals of the rectangle $MNPQ$ are perpendicular to each other. But this property of perpendicularity of the diagonals is a property which distinguishes squares from other rectangles.

Our proof is completed.

[1] We hope that no reader will feel that there is a contradiction in the combination of words "A rectangle . . . is . . . a square." Of course, not all rectangles are squares, but some of them are.

EXAMPLE 10. *Two lines, exactly one of which is perpendicular to a third line, do not intersect.*

This is a variation of an ancient sophism which has come down to us from the Greek mathematician Proclus (5th century A.D.).[1] Let us state the content of our assertion more precisely: At the points $A$ and $B$ of a straight line $AB$ (Fig. 14) draw two half-lines lying on the same side of the line $AB$ (in order to stress that these are in fact half-lines, they are marked with arrows in the drawing). Let $AQ$ form an acute angle $BAQ$ with $AB$, and $BP$ be perpendicular to $AB$; we shall prove that these half-lines do not intersect.

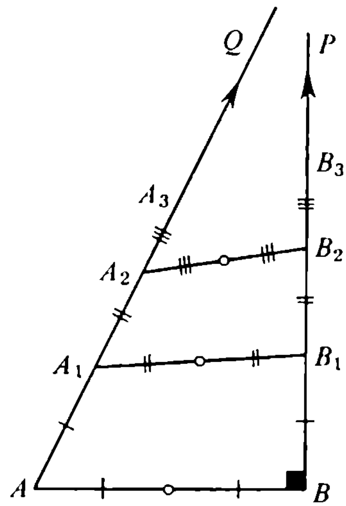



Fig. 14

Bisect the line segment $AB$ and on each of the half-lines $AQ$ and $BP$ lay off $\frac{1}{2}AB$; thus, $AA_1 = BB_1 = \frac{1}{2}AB$. The perpendicular half-line and the oblique half-line cannot intersect anywhere along $AA_1$ and $BB_1$; that is, the segments $AA_1$ and $BB_1$ cannot have a common point. In fact, if a common point $K$ did exist, we would obtain a triangle $AKB$ in which the sum of two sides $AK + KB$ is less than or equal to the length of the third side $AB$, and that is not possible. Joining the points $A_1$ and $B_1$, we repeat the previous construction; on each of the half-lines $AQ$ and $BP$ and in the direction of the half-lines, lay off segments equal to $\frac{1}{2}A_1B_1$ from the points $A_1$ and $B_1$. We thereby obtain $A_1A_2 = B_1B_2 = \frac{1}{2}A_1B_1$. By the reasons advanced above, the line segments $A_1A_2$ and $B_1B_2$ cannot have a common point, and, in particular, $A_2$ cannot coincide with $B_2$. That being the case, we bisect the line $A_2B_2$ and lay off $A_2A_3 = B_2B_3 = \frac{1}{2}A_2B_2$, and so on. It must be stressed particularly that the equal distances $A_nA_{n+1} = \frac{1}{2}A_nB_n$ are laid off in the direction of the half-line in question each time. The process will continue indefinitely. It could be stopped if the segment $A_nB_n$ were to disappear, that is, if for some number $n$ the points $A_n$ and $B_n$ were to coincide; but, as we have seen, that is impossible. Besides, the impossibility of such a coincidence is clear directly from the fact that we would then obtain a right triangle for which the

[1] For an exposition of this sophism according to Proclus, see R. Bonola, *Non-Euclidean Geometry* (New York: Dover, 1955), pp. 5–6.

hypotenuse $AA_n$ is equal to the leg $BB_n$. Thus, at no step of this infinite process does the intersection of the perpendicular half-line with the oblique half-line occur; therefore, it does not occur at all.

We have just seen a series of arguments which at times seem no less convincing than a proof from a geometry textbook. Sometimes the reasoning was directed at proving obvious absurdities; at other times the falsity of what was being proved was not immediately apparent, but the reader was told beforehand that each example did contain a mistake. The time has now come to uncover these mistakes.

Before proceeding to Chapter 2 for the analysis of all the examples given so far, we urge the reader to attempt to find the mistake in each example independently. It may be that not everyone will always successfully complete every example; nevertheless, one's own reflection about any of the examples prepares the ground for reading the analysis of the example in Chapter 2. If, on the other hand, the reader is successful, he will probably want to compare his interpretation with that given in Chapter 2. In suggesting this work, in which most readers will be inexperienced, I think it useful to give some preliminary hints and advice.

1. To refute an invalid geometric proof means to find a logical mistake in it. The difficulty lies in the fact that such a proof is correct almost everywhere, but always contains a gap at some point, and that gap has to be discovered.

2. In criticizing a proof it is often pointed out that it is carried out on the basis of "an incorrect drawing." This is not a very good criticism; in any case, it is not possible to limit oneself to this. When we say that drawing $A$ is incorrect and should be replaced by drawing $B$, we often mask the following state of affairs: Not all possible cases are taken into consideration in the proof (and *that* is the logical mistake!). Conclusions valid for those cases which are considered and depicted in drawing $A$ *alone* may turn out to be invalid for the other cases (drawing $B$). The source of the mistake is thus not in the drawing, but in the incomplete determination of the possible cases.

3. If the case depicted in drawing $A$ leads to an absurd conclusion, it is sufficient to show that such a result is not obtained on the basis of drawing $B$, in order to prove indirectly the impossibility of the case $A$. It is also desirable, but not indispensable, to obtain

a direct proof of the fact that the hypotheses of the theorem lead necessarily to the case $B$. Examples of such proofs appear in Chapter 2.

4. Although a drawing cannot in itself reveal either the correctness or the falsity of an assertion, it is nevertheless recommended that all drawings be made as accurate as possible (by means of instruments). Where we are dealing with an obvious sophism, it is useful to make a drawing which will stress sharply the absurdity of the conclusion, for instance, in Example 7 by depicting an obtuse angle which is close to 180°, in Example 10 by drawing the two half-lines in such a way that they intersect within the limits of the drawing, and so on. Such a drawing can provide a clue for finding the mistake.

5. In some cases the mistake is in no way connected with the drawing, but consists, for instance, in that the proof is given (correctly) not for the assertion which one has set out to prove, but for an assertion related to it. Here the author of the proof has either not noticed the substitution himself, or else counts on others' not noticing it.

6. If it is not known whether or not the proposition being proved is true, it is best, though not obligatory, to begin by clearing up this question. It should be kept in mind that the assertion will have been refuted if even one example is constructed which contradicts it.

The reader will understand more clearly the significance of these hints after he has carried out independently the work suggested and read the succeeding chapter. I therefore recommend referring to these hints while reading Chapter 2 and considering them again after reading that chapter.

# 2. Analysis of the Examples Given in Chapter 1

EXAMPLE 1. In asserting that it is possible to rearrange the parts *I*, *II*, *III*, and *IV* of the square to obtain the rectangle we rely upon what seems to be obvious, or upon the evidence of a crude experiment, if we cut the pieces out of paper. On what basis can we conclude that a triangle is formed if the figures *I* and *III* (or, what amounts to the same thing, *II* and *IV*) are placed side by side, that is, that the oblique lateral side of trapezoid *I* and the hypotenuse of triangle *III* form a continuous straight line which is not bent at their common point? Of course the fact that we are not able to see such a bend in a drawing, or that we do not observe it when we perform the cutting experiment cannot be used as an argument. Even aside from the imperfection of our visual impressions, we note that they have to do not with geometric figures but with physical models of these figures, and therefore are of no use for rigorous geometric proofs.[1]

In order to recognize that the whole proof is unsound, it is sufficient to discover this gap. We may even refuse to discuss the matter further until this gap has been filled. However, we shall not follow this course but shall instead attempt to clear up the question of the bend completely.

If we could prove, for instance, that the angles $\alpha$ and $\beta$ in Fig. 1 are supplementary, or else that the angles $\alpha$ and $\alpha'$ (in the same figure) are equal, the absence of a bend would be confirmed and the validity of the proof established. Is this possible? Reasoning indirectly, the answer is negative, for a positive answer to this question would lead to the equality $441 = 442$.

However, it is possible to ascertain directly that the angles $\alpha$ and $\alpha'$ are not equal, and at the same time to determine which of

[1] In the history of mankind this was by no means understood at once. In excavations of an ancient Indian temple dating back to 1,000 B.C., some mathematical records were found, among them a geometric figure depicted on the temple wall. The drawing apparently had to do with a rule for finding the area of a circle; instead of a proof, the word "Behold!" was written alongside the figure.

them is the greater. The next few lines will be within the grasp of the reader if he knows even a little trigonometry. (Instead of trigonometry the properties of similar triangles could be employed.) From triangle *III* in Fig. 1 we find the tangent of the angle $\alpha$:

$$\tan \alpha = \tfrac{21}{8}.$$

If in trapezoid *I* we drop the perpendicular (not shown in Fig. 1) from the vertex of the angle $\beta$ to the longer base, a right triangle is formed with legs of length 13 and $13 - 8 = 5$. From this we obtain

$$\tan \alpha' = \tfrac{13}{5}.$$

But $\frac{21}{8} - \frac{13}{5} = \frac{1}{40}$; therefore, $\frac{21}{8} > \frac{13}{5}$, and $\tan \alpha > \tan \alpha'$. From this it follows that

$$\alpha > \alpha', \quad \alpha + \beta > 180°.$$

Now the picture is clear; the parts *I*, *II*, *III*, *IV* of the square may actually be placed into the rectangle, but they do not cover it completely; a gap remains in the form of a very thin parallelogram —a "crack" which lies along the diagonal of the rectangle. It is not surprising that we do not notice this crack; for over the total length of 36.4 . . . (cm.) it has an area of only 1 (cm.$^2$), which is exactly the excess area of the rectangle $R$ over the square $Q$. A reader who wishes to make the picture still more obvious might change the numerical data in Fig. 1, for instance, in the way shown in Fig. 15,

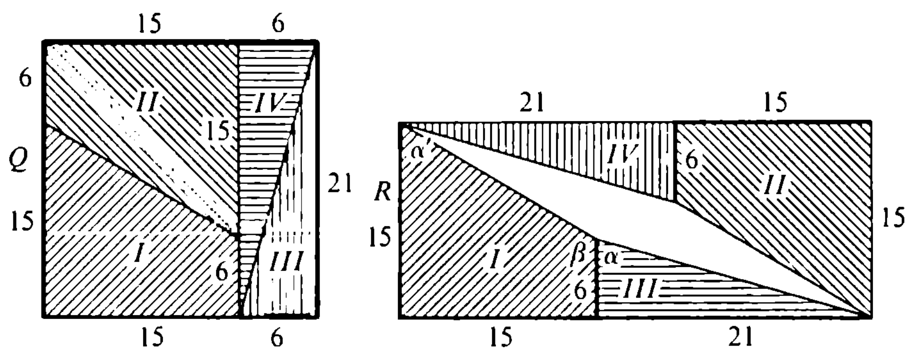



Fig. 15

where the "crack" has an area of 99 (cm.$^2$), and the entire rectangle a total area of 540 (cm.$^2$).

Example 2. The mistake made here is of a type which was well known in classical logic, bearing the complicated Latin name *ignoratio elenchi*, which freely translated means "failure to understand what has been proved," or proving something other than what was required.

What, indeed, does the argument using Fig. 2 actually establish? It is shown only that a unique straight line is obtained if the parallel line is constructed by the method described there (by means of two perpendiculars). But are there not other ways to construct parallel lines? Yes, it is well known that other constructions exist which lead to the same end.

For instance, instead of base $D$ of the perpendicular $CD$ from $C$ (see Fig. 2), we might take any other point $D'$ on the straight line $AB$ (Fig. 16), join it to $C$ by the straight line $D'F$, and on the

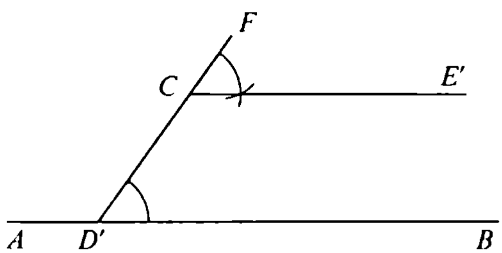


Fig. 16

half-line $CF$ at the point $C$ construct the angle $FCE'$ equal to the angle $CD'B$ in such a way that the half-lines $CE'$ and $D'B$ lie on the same side of $FD'$. On the basis of the theorem that straight lines making equal corresponding angles with a transversal are parallel, which can be proved without introducing the parallel postulate, it is possible to assert that the straight line $CE'$ is parallel to $AB$. But where is the guarantee that the straight line $CE$ in Fig. 2 coincides with $CE'$ in Fig. 16? To assert that different constructions lead to the same straight line is to accept without proof that which we set out to prove.[1]

[1] In the geometry of Lobachevskii the straight lines $CE$ and $CE'$ do not coincide. Interested readers may wish to refer to N. Lobachevskii's *Geometrical Researches on the Theory of Parallels*, trans. by George B. Halsted, Open Court Publishing Co.

EXAMPLE 3. In considering just the three cases, we have made the following assumption: For any pair of parallel lines intersected by any third line, the sum of any pair of interior angles lying on one side of the third line will be always greater than, always equal to, or always less than 180°. Only at first glance will it appear that all possible cases have been exhausted; the possibility that the sum of the interior angles on one side of the third line is sometimes greater, sometimes less, and sometimes equal to 180° has been omitted. This assumption does not lead to any contradiction. For instance, the supposition that

$$\angle 1 + \angle 4 > 180^\circ \text{ and } \angle 2 + \angle 3 < 180^\circ$$

(see Fig. 3) may not in any way contradict the relationship

$$\angle 1 + \angle 2 + \angle 3 + \angle 4 = 360^\circ.$$

Note that without going into a detailed analysis of the proof it is possible to discover its lack of substance at the very outset by one simple observation—the proof does not use the fact that the straight lines $AB$ and $CD$ are parallel. If the proof were correct as it stands, the following theorem would have been proved: "If any pair of straight lines is intersected by a third line, the sum of the interior angles on one side of the third line is equal to 180°." But this is incorrect. It is precisely when the parallelism of the straight lines $AB$ and $CD$ is discarded that, as a rule, the fourth possibility, which was omitted in the fallacious proof, will materialize—the sum of the interior angles on one side of the intersecting line will be greater than 180°, and on the other side less than 180°.

EXAMPLE 4. We have become accustomed to accept the statement that the sum of the angles of a triangle is constant (180°) for all triangles, irrespective of their shape and dimensions. For this reason the majority of us do not protest against the statement, "We shall denote the sum of the angles of any triangle by '$x$'." But actually when we set out to prove the theorem in question, nothing is known concerning the sum of the angles of the triangle, and there is no basis whatsoever for assuming that it is the same for all triangles. Of course, we could accept without proof the fact that the sum is the same and in that case the arguments advanced would indeed prove that this sum is equal to 180°. But this would merely mean that we had introduced another postulate in place of the parallel postulate.

EXAMPLE 5. In the history of mathematics several cases are known in which the very same mistake has been made; that is, with no basis it has been accepted that the terms of a given infinite set must necessarily include a greatest term (or a least term).

Now, it would not occur to anyone to seek the greatest term of the numbers

$$1, 2, 3, \ldots$$

of the sequence of natural numbers, as the absence of such a number is explained by the fact that in this case the numbers are increasing all the time and this sequence has no end. Moreover, the sequence of fractions whose numerator is one less than the denominator,

$$\tfrac{1}{2}, \tfrac{2}{3}, \tfrac{3}{4}, \tfrac{4}{5}, \ldots,$$

may also be continued indefinitely by repeatedly adding 1 to both the numerator and the denominator. As in the case of the first sequence, the numbers will increase, and there is no greatest term among them.

Closer to our problem is the following example taken from geometry: The interior angle of a regular polygon, equal to

$$\left[\frac{180(n-2)}{n}\right]^\circ$$

where $n$ is the number of sides, is always less than 180°, but there is no regular polygon with a greatest interior angle.

The weak point in the proof under consideration is precisely the assumption that among all triangles whose angles we know do not add up to more than 180°, there exists a triangle for which this sum has a greatest value. This is an unproven assertion which we could accept as a new postulate in place of the parallel postulate.

By combining the results of the analysis of Examples 4 and 5 we conclude that it is possible to prove that the sum of the angles of a triangle is equal to 180° and at the same time render the parallel postulate superfluous if we accept without proof one of two assertions:

(1) The sum of the angles of all triangles is the same.

(2) There exists at least one triangle for which the sum of the angles is greatest.

EXAMPLE 6. Not all possible cases have been examined (in this connection it is useful to recall the footnote on page 7): in fact, no account has been taken of the possibility that one of the two perpendiculars $NP$ and $NQ$ falls on a side of the triangle $ABC$ while the other falls on the extension of a side. (See Fig. 17, ignoring the

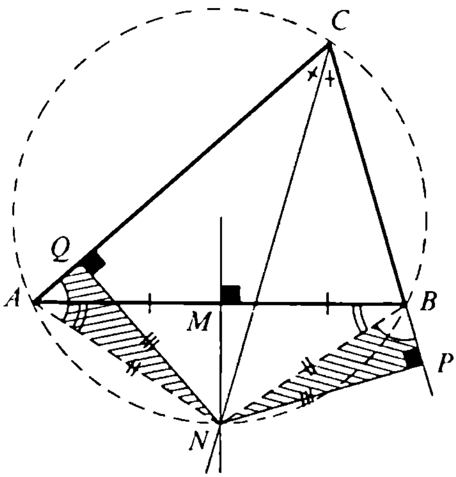



Fig. 17

circle for the present.) If this occurs, one of the angles at the base $AB$ of the triangle $ABC$ will be found to be the difference of two angles, while the other will be adjacent to the sum of the same two angles; from this, of course, no conclusion can be drawn regarding the equality of the angles at the base and of the lateral sides $AC$ and $BC$. Establishing this gap in the proof is sufficient to discredit it. Moreover, if the given triangle is *not* isosceles, it may be asserted, arguing indirectly, that none of the cases considered (Fig. 5, 6, 7) will occur and that the only possible case (Fig. 17) has been omitted.[1]

We shall now give a direct proof that the parts of a nonisosceles tri-

[1] At first it may seem that the case for which the point $N$ lies inside or on the base of the triangle and the points $P$ and $Q$ lie on different sides of $AB$ has also been omitted. It is quite possible that a perpendicular dropped from a point inside the triangle to one of its sides should fall on the extension of the side; it is sufficient to consider an obtuse triangle. In the next paragraph we shall, however, establish that for a nonisosceles triangle the point of intersection of the perpendicular bisector of the base and the bisector of the opposite angle *must* lie outside the triangle. If the reader is acquainted with the theorem that the bisector of an angle of a triangle divides the opposite side into parts proportional to the two remaining sides, he might attempt to establish this property of the point of intersection by a different method.

angle are arranged precisely as shown in Fig. 17. For convenience, we suppose that $CA > CB$. We circumscribe a circle about the triangle $ABC$. From the property of inscribed angles, the bisector of angle $C$ must pass through the mid-point $N$ of the arc $AB$ which angle $C$ intercepts. But the perpendicular bisector of the chord $AB$ must pass through the same mid-point $N$. Thus, the point of intersection of the bisector of $\angle C$ and the perpendicular bisector of $AB$ falls on the circumscribed circle; that is, it lies *outside* the triangle $ABC$.

The perpendiculars dropped from $N$ to sides $CB$ and $CA$ will fall on these sides or their extensions, depending on whether the angles $NAC$ and $NBC$ are acute or obtuse. Instead of these inscribed angles we shall examine the arcs which they intercept. Since we assumed $CA > CB$, we have $\overparen{CA} > \overparen{CB}$, and from $\overparen{AN} = \overparen{BN}$ it follows that $\overparen{CAN} > \overparen{CBN}$. This means that $\overparen{CAN}$ is greater than a semicircle and $\overparen{CBN}$ is less than a semicircle. Consequently, the angle $CBN$ is obtuse and the angle $CAN$ acute. The perpendicular $NP$, therefore, falls on the extension of $CB$, while the perpendicular $NQ$ falls on the side $AC$ itself. (As an exercise we suggest that the reader prove that the points $P$, $M$, and $Q$ lie on a straight line.)

EXAMPLE 7. The proof appears at first to be convincing, as it creates the illusion that all essentially different cases have been examined[1]—the point $N$ lies above, below, or on the straight line $AB$. However, the course of the proof does not depend *only* on the position of the point $N$. Notice that in case (3) the right angle $ABD$ together with the acute angle $ABN$ will always give an obtuse angle $DBN$; however, the obtuse angle $CAB$, when added to the acute angle $NAB$, may either again give an obtuse angle (Fig. 9) or else a reflex angle (greater than 180°, see Fig. 18), which changes the matter fundamentally.

Thus, case (3) must be divided into two subcases: the obtuse angle $CAB$ and the triangle $CAN$ lie (1) on the same side of the straight line $AC$ (see Fig. 9), or (2) on opposite sides of it (see Fig. 18, which should first be examined ignoring the dotted lines). The first subcase, in which the angle $CAB$ is a part of the angle $CAN$, has been examined and leads to the equality of the angles $DBA$ and $CAB$. The second subcase, however, does not lead to this

[1] Two cases must be regarded as essentially different if a proof which is valid for the one case cannot be applied word for word to the other case.

result; as before, the right angle $DBA$ is the *difference* of two angles ($\angle DBN$ and $\angle ABN$), but the obtuse angle $CAB$ together with the

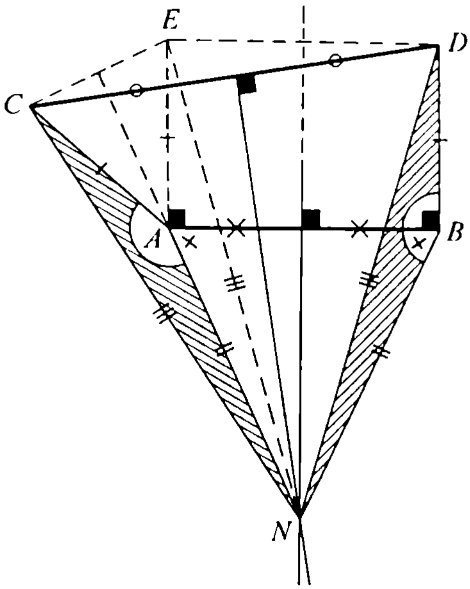



Fig. 18.

*sum* of the corresponding angles ($\angle CAN$ and $\angle BAN$) totals 360°. Arguing indirectly we can conclude that the second subcase is the only one possible.

We shall carry out an additional construction which will afford a clearer view of the arrangement of the parts of the figure. At the point $A$ erect a perpendicular to $AB$ (now the dotted lines in Fig. 18 come into the picture), and on it lay off the line segment $AE$ equal and parallel to $BD$; evidently, we have $AE = AC$. Join $E$ to the points $D$, $N$, and $C$. Since $ABDE$ is a rectangle, the perpendicular bisector of $AB$ will also be the perpendicular bisector of $ED$; consequently, $NE = ND$, and hence $NE = NC$. Thus, each of the points $A$ and $N$ is equidistant from the end points of the line segment $CE$; consequently, the straight line $AN$ is the perpendicular bisector of this line segment. The triangle $DBN$ is transformed into the triangle $EAN$ (oriented in the opposite direction) by a reflection in the perpendicular bisector of $AB$, and the triangle $EAN$ is transformed in turn into the triangle $CAN$, with the same orientation as the triangle $DBN$, after a reflection in the straight line $AN$. Thus, $CAN$ can be obtained from $DBN$ simply by a rotation about the vertex $N$ through $\angle BNA$, which is equal to $\angle EAC$, that is, the difference between the original obtuse and right angles.

EXAMPLE 8. Let us make certain first of all that the theorem is not true. For this purpose it is sufficient to furnish a counterexample, that is, a case in which the conditions of the theorem are fulfilled, but not the conclusion. We obtain such an example if we divide any isosceles triangle $LMN$ ($LN = MN$, Fig. 19) into two parts by means

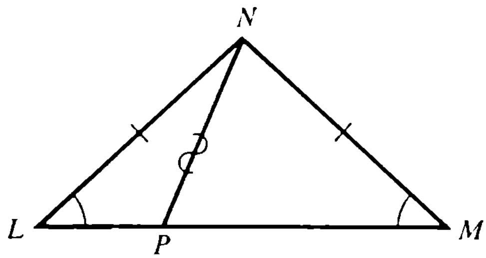



Fig. 19

of an oblique line segment $NP$ from the vertex. In the triangles $LNP$ and $MNP$ so obtained, $NP$ is a common side, and, moreover, $LN = MN$ and $\angle L = \angle M$. Thus, the conditions of the theorem are fulfilled, although the two triangles are, of course, not congruent as $LP \neq MP$.

But even if it is not known whether or not the theorem is true, it is possible to discover a gap in the proof. This gap lies in our omission of the cases in which the straight line $CC_2$ passes through one of the end points of the line segment $AB$, that is, in which the sides $CA$ and $C_2A$ or $CB$ and $C_2B$ are extensions of each other (using the notation of Figs. 10–12).

If the two *equal* sides $AC$ and $AC_2$ lie on the same straight line (Fig. 20), the conclusion of the theorem is still correct; by joining

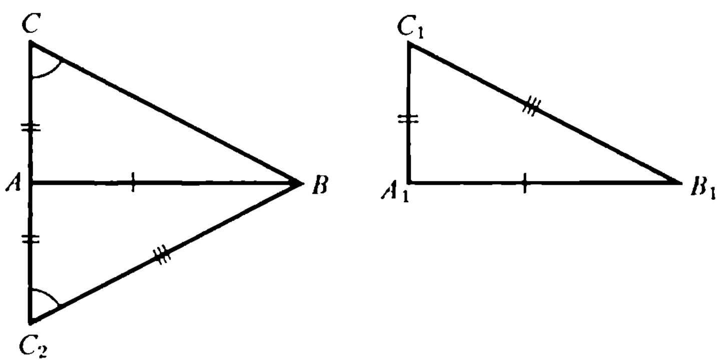



Fig. 20

the triangle $A_1B_1C_1$ to the triangle $ABC$, we obtain the triangle $BCC_2$, which is isosceles because angles $C$ and $C_2$ are equal; consequently, $BC = BC_2 = B_1C_1$, and the triangles are congruent.

Let us add that this occurs only with right triangles; in the left of Fig. 20 the angles at the point $A$ are equal and supplementary and, hence, right angles.[1]

A different picture is obtained if those sides, $BC$ and $BC_2$, about which nothing is known from the hypotheses lie on a straight line (Fig. 21). Of course, we obtain the isosceles triangle $ACC_2$, but it is

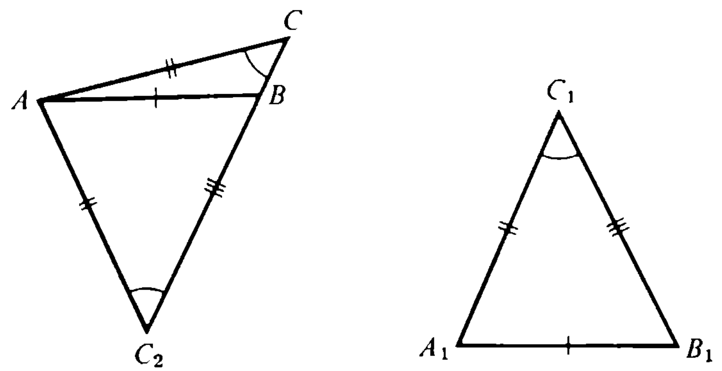



Fig. 21

impossible to draw from this any conclusions relating to the sides $CB$ and $C_2B$. Furthermore, the reader will immediately recall the figure of an isosceles triangle (Fig. 19) divided into two unequal parts. Hence, we can conclude that the triangles $ABC$ and $A_1B_1C_1$ are not congruent in this case unless the angles at the vertices $B$ and $B_1$ are right angles.

*Note 1.* The foregoing arguments suggest ways in which our theorem might be "amended," that is, how it might be replaced by some related theorems which are valid. We offer two "amended" versions.

1. *If two sides and the angle opposite one of them in one triangle are equal to the corresponding parts of another triangle, then the angles* ($B$ *and* $B_1$) *which lie opposite the other pair of equal sides will either be equal to each other and the triangles will be congruent* (*Fig.* 10–12, 20), *or else these angles will be supplementary* (*Fig.* 21), *and the triangles will not be congruent.*

2. *If two sides and the angle lying opposite the larger of them in*

[1] Note that this possibility of $C$, $A$, and $C_2$ lying on a straight line must also be anticipated in the usual proof for the congruence of triangles by three equal corresponding pairs of sides. There the corresponding sections of the proof can be carried out successfully, and it is thereby also found that the congruent triangles are right triangles.

*one triangle are equal to the corresponding parts of another triangle, then these triangles are congruent.*

The second statement excludes the case depicted in Fig. 21. Neither of the angles $B$ and $B_1$ can be obtuse or even a right angle, as $C$ or $C_1$ would not then lie opposite the larger side.

*Note 2.* The reader may at first be puzzled by the following assertion: The case of congruence of the two triangles can, in a certain sense, be made "as close as we like" to the case where they are not congruent. We may regard Fig. 21 as a "degenerate" case of Fig. 10, in which the point $B$ has slid onto the line $CC_2$. In Fig. 10, $B$ could lie as close as we like to $CC_2$, and $\triangle CBC_2$ could be as flattened as we like, but it would necessarily be isosceles, and the proof of congruence would be valid. But as soon as $B$ gets *onto* $CC_2$, the equality of $CB$ and $C_2B$ can no longer be demonstrated, and, indeed, may not hold. We shall discuss briefly a point of view which will throw some light on this.

It is sometimes convenient to regard three points which lie on a straight line as the vertices of a "degenerate" triangle: if $Q$ lies between $P$ and $R$, the angles of the "triangle" will be $\angle P = 0°$, $\angle R = 0°$, $\angle Q = 180°$. The meaning of this terminology is clear; as long as the points lie "almost" on a straight line, they still determine a triangle with two extremely small angles, and the third angle is nearly an angle of 180°. The figure can be continuously deformed so that these three points come to be situated exactly on a straight line, and it is desirable to retain the previous terminology. Some theorems are valid for both "real" and degenerate triangles; for instance, the theorem that the sum of the angles of a triangle is equal to two right angles. There are, however, other theorems which are not valid for degenerate triangles. Among these, in particular, is the theorem stating that a triangle with two equal angles is isosceles. This theorem is valid if the equal angles differ from zero; but in the degenerate triangle $PQR$ described above, it does not follow at all from the equality $\angle P = \angle R = 0°$ that $PQ = QR$, that is, that $Q$ is the mid-point of $PR$; $Q$ might lie anywhere on the line segment. This example is directly related to Example 8. As long as the point $B$ in Fig. 10 or 11 does not lie on the straight line $CC_2$, then no matter how small the angles are at the side $CC_2$ in the triangle $BCC_2$, it follows from the equality of these angles that $CB = C_2B$. But if, as in Fig. 21, the triangle $BCC_2$ is degenerate, then both of the angles become

zero, and the conclusion that the sides are equal and with it the conclusion of the theorem, lose their validity.

EXAMPLE 9. The assertion of the theorem is erroneous, as we may show by constructing a counterexample. It is sufficient to take the sides of the rectangle parallel to the diagonals of the square, and its vertices not bisecting the sides of the square. More precisely, from two opposite vertices, $A$ and $C$, of the square lay off along its sides four equal segments $AM = AQ = CN = CP$ of any length not equal to $\frac{a}{2}$, where $a$ is the length of a side of the square. Then $MNPQ$ is a rectangle, for its sides make angles of 45° with the sides of the square, and so are parallel to its diagonals. But these are perpendicular, and so the sides of $MNPQ$ are also perpendicular (Fig. 22).

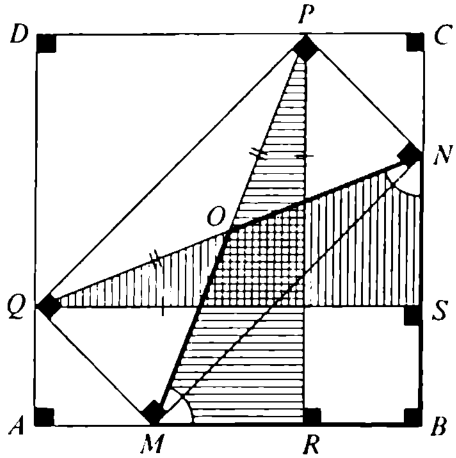


Fig. 22

Quite apart from this construction, we may see directly that the proof we gave in Example 9 was fallacious. Relying only on the appearance of Fig. 13, we assumed that of the two projections, $P$ onto $AB$ and $Q$ onto $BC$, one lies inside and the other outside the quadrilaterial $MBNO$. We see from Fig. 22 (where we do not now assume any more than is marked on the figure) that this may not be so. Our original proof, with $S$ and $R$ both lying *inside* $MBNO$, would show merely that $\angle OMB = \angle ONB$. Let us carry the argument further. $\triangle MON$ is isosceles, being formed by the diagonals of a rectangle. So $\angle OMN = \angle ONM$, and, by subtraction, $\angle BMN = \angle BNM$. By considering $\triangle BMN$, we see that these angles are 45°, and so our rectangle is exactly of the type we constructed above. To complete the analysis, we must examine the case where both $R$ and $S$ fall outside $OMBN$, which turns out to lead to the same result as the case we have just considered, and the case where one of $R$ and $S$ coincide with a vertex of $OMBN$, which turns out to give a square. We leave the proofs to the reader.

To summarize, we may replace the false theorem by a more complicated theorem such as:

1. *If a rectangle is inscribed in a square in such a way that one of the sides of the rectangle is not parallel to either of the diagonals of the square, then the rectangle is a square;* or

2. *If a rectangle with unequal sides is inscribed in a square, the sides of the rectangle must be parallel to the diagonals of the square.*

EXAMPLE 10. The logical mistake is the same as that in Example 2: "failure to understand what has been proved." In other words, for the proposition to be proved we have substituted another proposition which actually holds, but from which the statement to be proved does not in any way follow. Let us consider once more the line of reasoning and simplify the task of discovering the mistake by replacing Fig. 14 by Fig. 23, in which the half-lines $AQ$

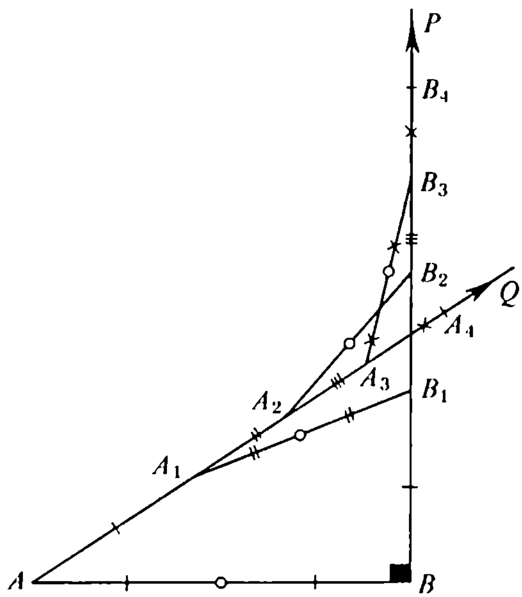



Fig. 23

and $BP$ do intersect. Both figures have been made without any deliberate distortion. Let us indicate by $AA_1$, $A_1A_2$, $A_2A_3$, ... the first, second, third, ... segments on the oblique line $AQ$, and let us indicate by $BB_1$, $B_1B_2$, $B_2B_3$, ... the first, second, third, ... segments on the perpendicular $BP$. It must be taken as proved that (1) the process of laying off these segments may be continued

indefinitely so that it is possible to obtain segments with arbitrarily high index, and that (2) segments with the *same index* do not intersect, that is, the first segment of the perpendicular has no common point with the first segment of the oblique line, the second segment of the perpendicular has no common point with the second segment of the oblique line, nor the hundredth with the hundredth, etc. But why should not segments with *different indices* intersect, say, the twentieth segment of the perpendicular with the twenty-fifth segment of the oblique line? For when we assert that the perpendicular and the oblique line do not intersect anywhere, we must prove that none of the segments of the perpendicular has a common point with *any* of the segments of the oblique line. And we must not be satisfied with proving instead that none of the segments of the perpendicular intersects the *corresponding* segment of the oblique line. If we turn to Fig. 23, in which the notation of Fig. 14 is preserved, we see by inspection that the second segment of the perpendicular and the fourth segment of the oblique line do intersect.[1] This sophism is noteworthy for the contrast between the elementary nature of the mistake and the difficulty of discovering it.

*Note.* As we have said before (see the footnote on page 16), we have used only the idea underlying the sophism given by Proclus. He examines two straight lines taken arbitrarily—in actual fact, two half-lines which do not lie on a straight line and which have different origins—and proves by means of the endless process of laying off segments described above, that these straight lines do not intersect. Proclus correctly characterizes the logical error contained in this sophistic argument when he says that the only thing proved is that the point of intersection cannot be found by the method of construction used, but this does not in any way signify that such a point does not exist. Judging from the account given by Bonola, however, it is not certain that Proclus penetrated more deeply into the geometric substance of the mistake; in any case the 19th-century Italian author is clearly mistaken when he states that the inaccessability of the point of intersection is brought about by the same reasons that give rise to the famous sophism about "Achilles and the tortoise." By this comparison Bonola means, of course, that the point of intersection, say $K$, of the half-lines $AQ$ and $BP$ is unattainable by the given construction simply because as $n$ increases without bound, the points $A_n$ and $B_n$ approach $K$ as their limit, but never reach it. In our version such an assumption is not possible, for from the equality $AA_n = BB_n$, which holds for arbitrary $n$, it would follow that $AK = BK$, that

[1] Knowing the angle $A$, it would be possible to calculate the indices of the intersecting segments by means of trigonometry.

is, that the hypotenuse is equal to one of the legs in triangle $ABK$. But this is impossible here and also in the construction of Proclus-Bonola except when the triangle $AKB$ is isosceles. Thus, what happens generally is that segments with different indices intersect, and segments with the same index do not tend to a common limit.

CONCLUSIONS. The reader may ask the following: If mistakes in mathematical arguments are sometimes masked to such an extent that they can be discovered only after careful analysis, does mathematics really provide such a reliable foundation for the exact sciences (physics, engineering, and others) as we customarily believe?

Of course, no scientific method is a guarantee against faulty conclusions; the method in question must also be applied correctly. This only shows that one should study the sources of possible errors and be more exacting in substantiating one's assertions. In order to perceive the danger of committing an error which may pass undetected, we must turn to the history of our science.

This history has witnessed various mistakes in the works of mathematicians, but these mistakes have never stopped the progress of the science and have been exposed when a higher stage of development was reached. As an imposing example we may cite the previously mentioned history of the attempts made through many centuries to prove the parallel postulate. About this postulate Lobachevskii wrote in 1823, "It has not been possible to find a rigorous proof for this truth thus far. Such proofs as have been given can be called only explanations, and do not deserve to be dignified as mathematical proofs in the full sense of the term." Lobachevskii arrived at this conviction a few years before his outstanding discovery. From the new level of this discovery in the history of geometry it became obvious, first to Lobachevskii and subsequently to the whole mathematical world, that the most ingenious plans for proving the parallel postulate could never succeed.

# 3. Mistakes in Reasoning Connected with the Concept of Limit

The examples in this chapter are within the grasp of readers with some acquaintance with the simplest properties of circles, the concept of a limit, trigonometry, and some solid geometry.

EXAMPLE 11. *The circumferences of all circles are of equal length.* This ancient sophism is ascribed to the Greek philosopher Aristotle (fourth century B.C.), and for a reason which will soon become clear it is called "Aristotle's wheel."

Let us recall arithmetic problems in which we are given the length of the circumference of a wheel of a cart or car moving along a road, and we are to find the distance traveled, or vice versa. As the basis for the solution we take the seemingly obvious fact that for each complete turn the rolling wheel travels a distance equal to (the length of) its circumference. If, for instance, the circumference of the wheel is 2 meters and in rolling it has made 30 complete revolutions, the distance through which it has traveled will be 60 meters. Where the motion takes place in a straight line, and where no special accuracy is required, these calculations can be confirmed by experiment. The circumference of the wheel may be measured with a tape; the completion of a revolution of the wheel may be judged by marking one of the spokes of the wheel, or instead of this by attaching to the rim a band which leaves an imprint on the ground. (Many metering devices installed on different transport vehicles register the number of revolutions but indicate the distance, or, in combination with a clockwork mechanism, the velocity.) Of course, all these calculations are correct in practice only if the wheel turns "normally," that is, if it does not jump or slip; in the language of mechanics this is expressed by the phrase "the wheel rolls without slipping."

We shall now return to our sophism. Let us examine two concentric circles $C$ and $C_1$ of different radii, which are rigidly attached to each other (Fig. 24). At the same time think of a physical model,

two cylindrical rollers mounted on a common axis, which we shall take as horizontal, and rigidly attached to each other. (We could do with a cylindrical roller, part of which has the form of another cylindrical roller with the same axis but with a smaller radius; see the picture in Fig. 24). Draw the tangents $MN$ and $M_1N_1$ to the

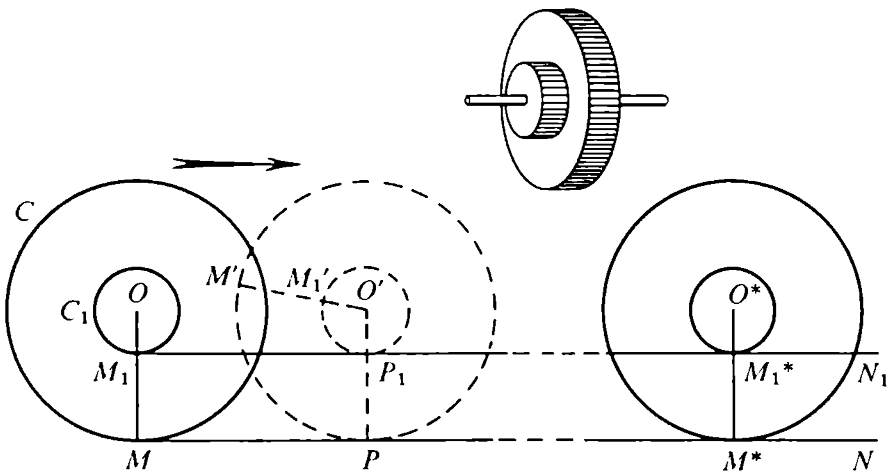



Fig. 24

circles $C$ and $C_1$ at the points $M$ and $M_1$, respectively, lying on the same radius $OM$. Since the circles are rigidly attached to each other, they will move as a unit; if one circle turns through a given angle, the other will turn through the same angle. Therefore, if the circle $C$ rolls along the straight line $MN$, the circle $C_1$ will roll along the straight line $M_1N_1$. In Fig. 24 the direction in which the circles roll is indicated by the arrow, and one of the intermediate positions is depicted by dotted lines, the points $M'$ and $M_1'$ being the new positions of the points $M$ and $M_1$. On the physical model we imagine a horizontal rail being placed under each of the cylindrical rollers; when the larger roller rolls on its rail it will cause the smaller one to roll on its own rail. Let the circle $C$, rolling along the straight line $MN$, make one complete revolution, as a result of which the point $M$ will take up the position $M^*$; the circle $C_1$ will then also have made a complete revolution, and the point $M_1$ will then occupy the position $M_1^*$ on the radius $O^*M^*$. $O^*M^*$ is parallel to $OM$ because both these radii are perpendicular to $MN$. We conclude that

$$MM^* = M_1M_1^*.$$

that is, that the two rolling circles go through identical distances in making a complete revolution, and that means that their circumferences are equal. As the circles $C$ and $C_1$ may be taken arbitrarily, we have furnished the required proof.

*Hint.* We shall not decide beforehand the question as to how the reader will overcome the evident contradiction obtained; the author's deliberations on this problem will be given in Chapter 4. The following observation would seem to be useful, whatever course is taken in pondering over this sophism.

The circle is often looked upon as the limit of a succession of regular polygons inscribed in it (or circumscribed about it) as the number of their sides increases without bound.[1] This suggests that in order to obtain a clearer picture of the process undergone by the rolling circle we might instead examine a regular polygon rolling on a straight line. The greater the number of its sides, the more closely will it approach the picture of a rolling circle.

The meaning of the statement "the (convex) polygon rolls without slipping along the straight line" seems obvious; we establish a certain order of passing around the sides of the polygon, say counterclockwise, and lay one of the sides on the straight line in the starting position. We rotate the polygon about that vertex which is common to this side and the side which follows it, until the second side lies on the straight line; we then rotate about the next vertex, and so on. In short, the polygon is "tipped" from one side onto the next one, rotating each time about the vertex which is common to these sides, and as a result of this it is transported along the straight line in the direction shown.

In the case of a regular $n$-gon ($n = 8$ in Fig. 25), label its vertices $A_1, A_1, \ldots, A_{n-1}, A_n$ and for the initial position let the side $A_1A_2$ lie on the straight line along which the polygon is to roll in

[1] The term "limit" is used here deliberately, instead of the widely used "limiting position," and it has a completely precise meaning: However narrow a ring between two circles concentric with a given circle, the radius of one of the concentric circles being larger and the radius of the other smaller than the radius of the given circle (for instance, the ring may be bounded by circles of radii $R - \varepsilon$ and $R + \varepsilon$, where $R$ is the radius of the given circle), it will be possible to find a number $n$ such that any inscribed (circumscribed) regular polygon with more than $n$ sides will lie completely within the ring. This must not be confused with the widely known proposition (more frequently, definition), "The length of the circumference of a circle is the limit of the sequence of lengths of the perimeters of regular inscribed (circumscribed) polygons, when . . .". As we shall see, the term "limit" has different meanings in the two cases (see below, Examples 12–14).

the direction $A_1A_2$. This direction is shown by the arrow in Fig. 25. In radian measure, the exterior angle of the polygon, and also the central angle, is equal to $\frac{2\pi}{n} = \frac{360°}{n}$. Therefore, it is sufficient to turn the polygon through the angle $\frac{2\pi}{n}$ about the vertex $A_2$ to make the side $A_2A_3$ lie on the straight line. After this rotation the

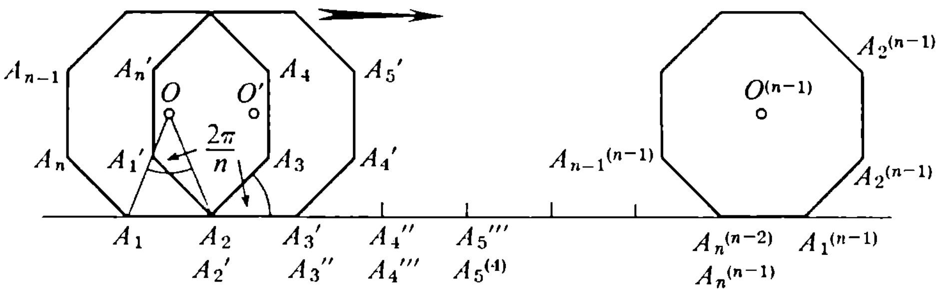



Fig. 25

center $O$ of the polygon occupies the position $O'$, and the vertices $A_1, A_2, A_3, \ldots, A_n$ take up the corresponding positions $A_1'$, $A_2'$ (coinciding with $A_2$), $A_3', \ldots, A_n'$. A further rotation about the vertex $A_3'$ through the angle $\frac{2\pi}{n}$ moves the polygon $A_1'A_2'A_3' \ldots A_n'$ to the position $A_1''\, A_2''\, A_3'' \ldots A_n''$; in Fig. 25 only the vertices $A_3''$ (coinciding with $A_3'$) and $A_4''$, both lying on the straight line, are marked. By continuing this process, we arrive after the $(n-1)$th rotation at the position $A_1^{(n-1)}A_2^{(n-1)} \ldots A_n^{(n-1)}$ with the center at the point $O^{(n-1)}$ and the side $A_n^{(n-1)}A_1^{(n-1)}$ on the straight line. As the vertex $A_1$ is then on the straight line once more, there is no need to continue the motion further; it is easy to see that the line segment $A_1A_1^{(n-1)}$ is equal to the perimeter of the polygon.

The reader will notice that the position of each vertex is marked by a pair of indices, the lower one showing the number of the vertex in the initial position, and the upper one, at first primes but subsequently numbers in parentheses, indicating the number of turns carried out; for instance, the symbol $A_6^{(4)}$ denotes the position of the vertex $A_6$ after the fourth turn. In addition to the properties inherent in the motion described in Fig. 25, it is possible to observe some accidental ones connected with the particular value chosen for $n$. It is suggested that the reader execute the drawing for some other, say odd, value of $n$, for instance, $n = 5$.

Turning once more to our sophism, we now take, instead of two concentric circles, two regular concentric $n$-gons whose corresponding sides are parallel, in other words, two regular polygons one of which is obtained from the other by a similarity (homothetic) transformation, with the center of similarity coinciding with the center of the second polygon. Assuming the polygons to be rigidly attached to each other, let us roll the bigger one along a straight line, as described above, and try to understand how the smaller polygon will then travel. Will the latter also be "tipped" from one side to the next? Will the perimeter of the smaller polygon be "unfolded" on the straight line as is the bigger one? (We could proceed in reverse order, rolling the smaller polygon and observing the motion of the larger one.)

PROBLEM. We shall now formulate a problem in which the rolling circle is replaced by a rolling polygon for a different purpose. This problem shares with the subject matter of the present chapter the property of passing to the limit in some instances which require justification (see the examples which follow).

It is known that when a circle rolls along a straight line, each point on that circle moves along a curve which is called a *cycloid*. If we trace the motion of the point which is the point of contact in the initial position of the rolling circle (Fig. 26, compare with Fig. 24), its path between two consecutive

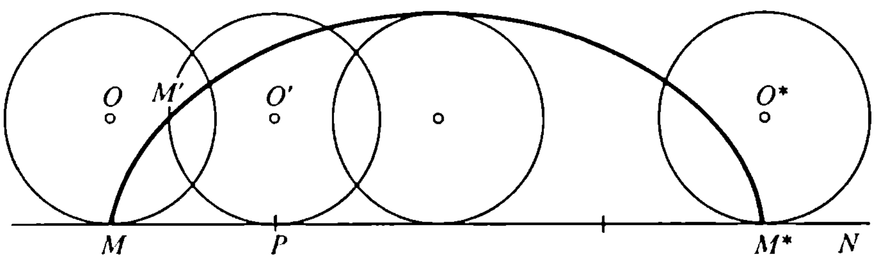



Fig. 26

points $M$ and $M^*$, which, as in Fig. 24, correspond to a complete revolution of the rolling circle, will have the form of the "cycloidal arc" $MM'M^*$. By means of higher mathematics it is established that the length of this arc is equal to eight times the radius of the rolling circle, and the area included between the arc and the straight line $MM^*$ is equal to three times the area of the circle. Our problem is to obtain these results by an elementary method. For this we follow the previous plan of replacing the rolling circle of radius $R$ by a regular $n$-gon inscribed in it.

In the notation of Fig. 25, the path of the point $A_1$ will be composed of $(n-1)$ circular arcs, $\overset{\frown}{A_1A_1'}$ with center $A_2$, $\overset{\frown}{A_1'A_1''}$ with center $A_3'$, ..., $\overset{\frown}{A_1^{(n-2)}A_1^{(n-1)}}$ with center $A_n^{(n-2)}$; these are not shown in Fig. 25. Together these circular arcs constitute a curve which goes from $A_1$ to $A_1^{(n-1)}$, resembling a cycloid but differing from it by the presence of "corners" where neighboring arcs meet. As the value of $n$ increases, the corners become smoother and the curve approaches the arc of a cycloid. It may be expected that the cycloid is the limit of this curve as $n \to \infty$. Now by means of elementary trigonometry it is easy to find the length of the path of a vertex of the polygon during one complete revolution for any value of $n$, since it consists of circular arcs; we can also find the area between that path and the straight line $A_1A_1^{(n-1)}$. If in the expressions obtained for the length and the area we pass to the limit as $n \to \infty$, we shall find the length to be $8R$ and the area to be $3\pi R^2$, respectively, which are the correct results.[1] However, this derivation of the formulas for the length and the area of the cycloid may be regarded as fully valid only after the passage to the limit has been justified, that is, after it has been proved that as $n \to \infty$, the values for the length and the area found for the rolling $n$-gon have as their limits the required length and area. It may be possible to do this within the realm of elementary mathematics, but it is certainly not easy.

The problem may be extended by also examining the paths of points inside or outside the circumference of a rolling circle and joined rigidly to it. We thereby arrive at so-called "prolate" and "curtate" cycloids. We can attempt to study these curves, too, by substituting for the rolling circle a regular inscribed polygon, and then passing to the limit.

[1] In the variant of the solution that the author has in mind, the following formulas not given in a secondary school trigonometry course are used:

$$\sin\alpha + \sin 2\alpha + \cdots + \sin k\alpha = \frac{\sin k\frac{\alpha}{2}\cdot\sin(k+1)\frac{\alpha}{2}}{\sin\frac{\alpha}{2}},$$

$$\sin^2\alpha + \sin^2 2\alpha + \cdots + \sin^2 k\alpha = \frac{2k+1}{4} - \frac{\sin(2k+1)\alpha}{4\sin\alpha},$$

$$2(\sin\alpha\sin 2\alpha + \sin 2\alpha\sin 3\alpha + \cdots + \sin(k-1)\alpha\sin k\alpha) = k\cos\alpha - \frac{\sin 2k\alpha}{2\sin\alpha}.$$

The reader can verify these identities, for instance, by induction on $k$. (For a study of this method see *The Method of Mathematical Induction* by I. S. Sominskii in this series.) Also involved is

$$\lim_{\omega\to 0}\frac{\sin\omega}{\omega} = 1 \qquad (\omega \text{ is the angle in radians}).$$

This formula may be found in many calculus books.

EXAMPLE 12. *The length of the hypotenuse of a triangle is equal to the sum of its two legs.*

In the right triangle $ABC$ from the mid-point $D$ of the hypotenuse we drop the perpendiculars $DE$ and $DF$ to the legs (Fig. 27, $C = 90°$); a broken line $BEDFA$ consisting of four sections is obtained, the length of which is evidently equal to the sum of the legs. We repeat this construction for each of the triangles $DBE$ and $ADF$; from the mid-point of the hypotenuses $DB$ and $AD$ we drop perpendiculars to the legs, thereby obtaining a broken line consisting of eight sections, whose length is the same as that of the preceding one. This process may be repeated an unlimited number of times: the hypotenuse will be successively divided into 2, 4, 8, 16, . . . equal parts; a series of saw-like broken lines appears—for the sake of brevity we shall refer to them simply as "saws"—joining the points $A$ and $B$ and consisting, respectively, of 2, 4, 8, 16, . . . "teeth" (that is, 4, 8, 16, 32, . . . sections). All of the saws are equal in length (that is, the sums of the lengths of the sections are equal), being the sum of the lengths of the legs. As the number of sections increases, the saw will approximate more and more closely the hypotenuse $AB$, so that for a very large number of sections it will be difficult in practice to differentiate between the broken line with the very small sections and the continuous straight line segment (just as it is difficult to distinguish between a circle and a regular polygon with a very great number of sides inscribed in it).

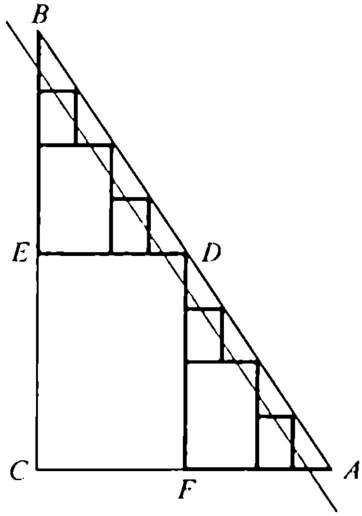


Fig. 27

We shall base a precise statement on this visual representation: The sequence of saws has the line segment $AB$ as its limit in the sense that the greatest distance between the points of the saw and the straight line tends to zero as we consider successive members of the sequence of saws. In fact this greatest distance is simply the altitude dropped to the hypotenuse of any one of the equal right

triangles which form the "teeth" of the saw, and the altitude of a tooth is less than its hypotenuse, which tends to zero. In other words, however narrow the strip between the hypotenuse $AB$ and a straight line parallel to it which intersects the legs (Fig. 27) in the sequence of saws there exists one which, together with all those succeeding it, will fit into this strip all the way from $A$ to $B$ (see the footnote on page 36). But the length of all saws is the same, which means that the sequence of their lengths consists of identical numbers and its limit will be that same number, which is equal to the sum of the legs. On the other hand, the hypotenuse is the limit of the saws; hence, its length must also be the limit of the sequence of their lengths. But as a sequence cannot have two different limits, our assertion is proved.

*Note 1.* It is not essential that the triangle $ABC$ be a right triangle. For an oblique triangle it is possible to construct a sequence of saws by drawing, through a point dividing one of its sides, lines parallel to the other two sides. Nor is it essential that the side be divided into 2, 4, 8, . . . equal parts; it may be divided into 2, 3, 4, 5, . . . or even into unequal parts, as long as their number increases without bound and as long as the length of the longest part tends to zero.

*Note 2.* The reader will perhaps look for the source of the mistake in the fact that the length of the saw remains unchanged, and that it would therefore appear to be impossible to speak of its limit. The rejoinder to this is that in mathematics we do consider sequences consisting of terms all equal to each other. According to the precise meaning of the concept "limit," this common number itself is the limit of such a sequence. However, it would not be difficult to modify our construction in such a way that the length of the saw becomes variable while everything else remains valid. It would be sufficient, for instance, to break off one of the teeth from each saw, let us say the first one counting from the point $A$; more strictly speaking, we replace the first two sections by a segment of the hypotenuse, beginning at $A$. This would decrease by one the number of sections of each saw. As before, the "damaged saws" will have the line segment $AB$ as their limit, but their lengths will each differ from the sum of the lengths of the legs, $AC + BC$, by a very small amount and will tend to this sum as a limit.

EXAMPLE 13. *The number $\pi$ is equal to* 2.

On the line segment $AB$ as diameter construct a semicircle (Fig. 28); then bisect the segment $AB$ and with each half as diameter

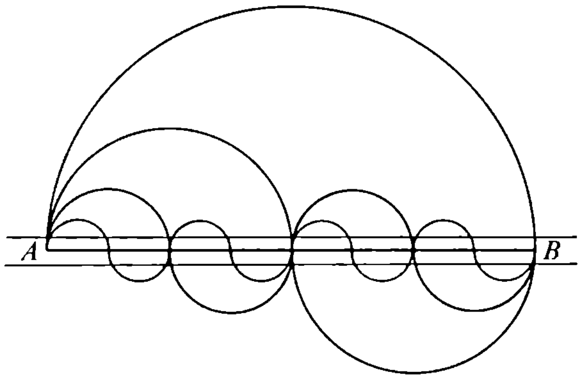



Fig. 28

construct semicircles which lie on different sides of $AB$. These two semicircles form a wavy line (resembling the sine curve), whose length from $A$ to $B$ is equal to the length of the original semicircle, that is, $\frac{\pi}{2}AB$. Each smaller semicircle is half as long as the larger one, since its diameter is half the length. Now divide the line segment $AB$ into four equal parts and construct a wavy line consisting of four semicircles (Fig. 28), the sum of whose lengths is again $\frac{\pi}{2}AB$. We repeat this process over and over again, dividing $AB$ into 8, 16, . . . equal parts and constructing on them semicircles which lie on alternate sides of the straight line $AB$. We obtain a series of wavy curves which approximate more and more closely the line segment $AB$. This segment represents their limit in the sense that the greatest of the distances of the points of each wavy line from the straight line $AB$ tends to zero as we proceed to successive members of the sequence; this greatest distance is evidently equal to the radius of the semicircles of which the line is composed. (In Fig. 28 a strip between two lines parallel to $AB$ is depicted. However narrow that strip, it is possible to find a place in our sequence from which all succeeding wavy lines will lie completely inside this strip from $A$ to $B$.) But the lengths of all these wavy lines are identical and equal to $\frac{\pi}{2}AB$; this must, therefore, also be the length of

the limit of these lines, that is, the length of the line segment $AB$. From the equality $\frac{\pi}{2} AB = AB$, we find that $\pi = 2$.

*Note.* This example may be supplemented by considerations analogous to notes 1 and 2 in Example 12; neither the manner in which the line segment $AB$ is divided nor the constant lengths of the wavy lines plays any essential part. As in the preceding example, in each wavy line it would be possible to replace one of the semicircles by its diameter, and then the length of the lines would be variable. We suggest that the reader examine another variant: Instead of semicircles, that is, arcs which are subtended by right angles, we may construct arcs which are subtended by an arbitrary different angle, constant or variable according to some given law, depending on the number of divisions, but not tending to 180°. We then obtain a different value for the number $\pi$.

EXAMPLE 14. "*Schwarz's cylinder.*"

To measure the length of the arc $\widehat{AB}$ of a curved line (Fig. 29, left), we can proceed almost in the same way as when measuring

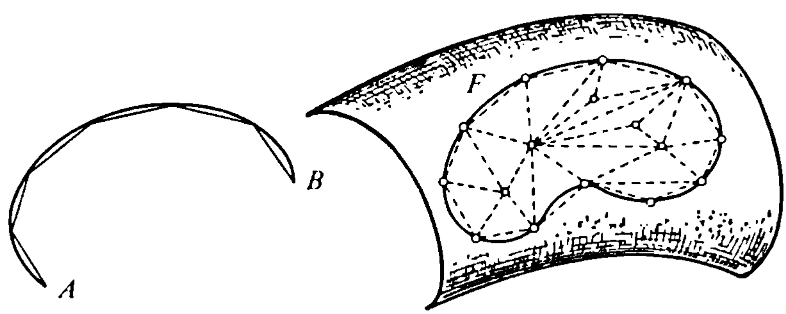



Fig. 29

a circumference or its parts. Broken lines are inscribed in the arc; the successive segments making up such a line will generally not make equal angles with each other or be equal in length. An infinite sequence of such broken lines is constructed, bringing the vertices infinitely close to each other, with the condition that as we consider successive members of the sequence, the length of the longest section of the broken line tends to zero. The limit of the sequence of lengths of these broken lines will then be the length of the original arc.

Passing on to two dimensions, we consider the analogous problem of finding the area of a figure $F$ which lies on a curved surface (Fig. 29, right). By analogy with the arc, it appears natural to proceed as follows: In the given figure inscribe polyhedral surfaces[1] with faces of decreasing size; the area of the figure $F$ will be the limit of the areas of these polyhedral surfaces, that is, of the sum of the areas of their faces. This is defined more precisely as follows: Inside the figure $F$ and on its boundary take a set of points and pass planes through groups of three points to obtain a polyhedral surface with triangular faces such that no two of these faces have any common interior points and no three of them have a common edge. Construct an infinite sequence of such polyhedral surfaces

$$F_1, F_2, \ldots, F_n, \ldots$$

inscribed in the figure $F$, such that the length of the longest edge occurring among the sides of all the triangular faces of the surface $F_n$ tends to zero as $n \to \infty$ and such that every point of the figure $F$ is the limit for some sequence of points chosen successively on $F_1, F_2, \ldots, F_n, \ldots$.

It now seems to be clear that we can choose each polyhedral surface in such a way that its outside boundaries are inscribed in the boundary of the figure $F$ (see Fig. 29). It seems almost self-evident that the sequence of areas of the polyhedral surfaces $F_1$, $F_2, \ldots, F_n, \ldots$ has a limit, namely, the area of the figure $F$. As in the preceding examples, for all practical purposes the curved surface and the polyhedral surface inscribed in it are indistinguishable when the faces of the latter become very small. At the end of the last century, however, the German mathematician G. A. Schwarz gave a simple example demonstrating that this "self-evidence" is deceptive. We shall now proceed to give an account of this example.

Consider a right circular cylinder of radius $R$ and altitude $H$ (Fig. 30); we shall attempt to determine the area of its lateral surface by the method set forth above. For this purpose let us divide the altitude of the cylinder into $n$ equals parts and through these points of division pass planes perpendicular to the axis. These

[1] A polyhedral surface is inscribed in a curved surface if all of its vertices lie on the curved surface.

intersect the lateral surface of the cylinder in $n - 1$ circles, which, together with the bases, divide the lateral surface into $n$ equal cylindrical bands. In one of these circles inscribe a regular $m$-gon and draw the generators of the cylinder which pass through its vertices. These generators divide each of the remaining circles into $m$ equal parts; the points of division form the vertices of regular $m$-gons inscribed in these circles. The segments of the generators together with the sides of the inscribed polygons form $mn$ identical rectangles whose vertices lie on the surface of the cylinder. One of them, $MNPQ$, is marked in Fig. 30.

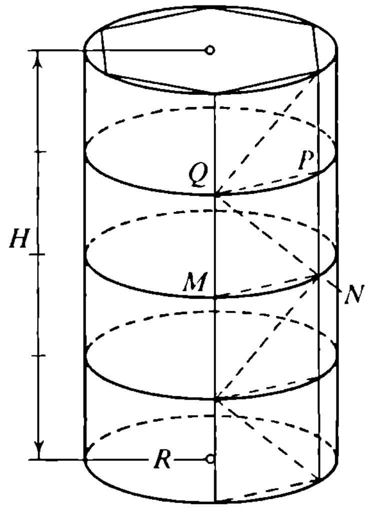


Fig. 30

Finally, dividing each rectangle into two triangles by means of a diagonal, we obtain a polyhedral surface consisting of $2mn$ identical triangular faces and inscribed in the lateral surface of the cylinder. When both $m$ and $n$ grow without bound, the sides of these faces tend to zero, and the distance of the points belonging to them from the lateral surface of the cylinder also tend to zero.[1]

Our inscribed polyhedral surface depends on the two indices $m$ and $n$. There are infinitely many ways of choosing a sequence of our polyhedral surfaces by making one of the indices a function of the other in such a way that both indices take on integral values and simultaneously tend to infinity. For instance, we could set $m = n$, or $n = 3m$, or $m = n^2$, etc. The reader has probably already noticed that our polyhedral surfaces actually coincide with the lateral surfaces of regular $m$-sided prisms inscribed in the cylinder. The purpose of dividing each rectangular face of these prisms into $2n$ triangles is only to make this method of inscribing the polyhedral surfaces conform to the general scheme depicted in Fig. 29. Thus, we have before us only a somewhat complex variation of the usual derivation of the formula for the area of the lateral surface of a cylinder ($S = 2\pi RH$), that is, inscribing a regular prism and taking the limit. Up to now our arguments do not contain any sophism.

[1] The distance of any point from the lateral surface of the cylinder is just the difference between the radius of the cylinder and the distance of the point from the axis of the cylinder.

We shall now alter somewhat the method of inscribing the polyhedral surfaces. As before, we divide the altitude $H$ into $n$ equal parts and draw our $n-1$ circular cross sections, which, together with the ends of the cylinder, give $n+1$ circles. In each of these circles we inscribe a regular $m$-gon, but with their vertices arranged differently, namely, in such a way that the generator drawn through any vertex of the polygon inscribed in one of the circles passes halfway between vertices of the polygon inscribed in adjacent circles. For instance, in Fig. 31 the generator through $P$ bisects the arcs $\widehat{MN}$ and $\widehat{QS}$; the figure does not show the straight lines $QM$ and $SN$ which are also generators. In other words, previously the polygon inscribed in any circle was obtained from the polygon inscribed in the neighboring circle simply by a parallel translation in the direction of a generator by a distance $\frac{H}{n}$, while now the translation is combined with a rotation about the center of the polygon through an angle equal to $\frac{\pi}{m}$. Arranging the regular inscribed polygons in this manner, we construct from the triangular faces a polyhedral surface (not convex!) by joining each vertex with the two vertices nearest to it on a neighboring circle. This polyhedral surface, which resembles a collapsible paper lantern in an extended position, consists of $2mn$ congruent isosceles triangles, $2m$ in each one of the $n$ strips; it is *inscribed* in the curved surface of the cylinder in the exact sense of that word (see the footnote on page 44).

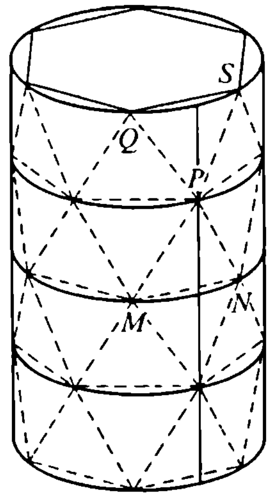



Fig. 31

To find the area of the polyhedral surface, examine one of its equal faces $MNP$, depicted in Fig. 31 and separately in Fig. 32 in enlarged form. Here $MN$ is the side of the regular $m$-gon inscribed in the circular cross-section with center $O$, the points $K$ and $L$ are the mid-points of the arc $\widehat{MN}$ and the chord $MN$, respectively, and $PK$ is a segment of the generator. Now $PM = PN$, since

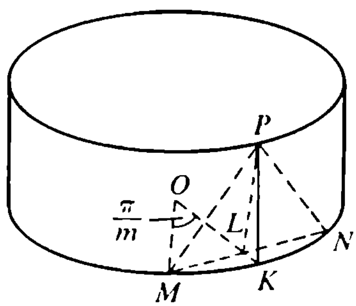



Fig. 32

these segments have equal projections onto the plane of the circle $O$; their projections are the equal chords $KM$ and $KN$. Therefore, triangle $MNP$ is isosceles. Its altitude $PL$ is found from the triangle $PKL$, in which

$$\angle K = 90^\circ, \qquad PK = \frac{H}{n},$$

$$KL = R - OL = R - R\cos\frac{\pi}{m} = 2R\sin^2\frac{\pi}{2m}.$$

Therefore,

$$PL = \sqrt{\left(\frac{H}{n}\right)^2 + 4R^2\sin^4\frac{\pi}{2m}}.$$

And since $\frac{1}{2} MN = R\sin\frac{\pi}{m}$, we have

$$\text{area } MNP = R\sin\frac{\pi}{m}\sqrt{\frac{H^2}{n^2} + 4R^2\sin^4\frac{\pi}{2m}}.$$

Letting $S_{m,n}$ be the area of the entire polyhedral surface obtained by dividing the circle into $m$ parts and the altitude into $n$ parts, we find that

$$S_{m,n} = 2mn\,R\sin\frac{\pi}{m}\sqrt{\frac{H^2}{n^2} + 4R^2\sin^4\frac{\pi}{2m}}$$

$$= 2mR\sin\frac{\pi}{m}\sqrt{H^2 + 4n^2R^2\sin^4\frac{\pi}{2m}}.$$

As already remarked in another connection, it is possible in an infinite number of ways to establish a relationship between the indices $m$ and $n$ and obtain a sequence from the areas $S_{m,n}$. Let us examine two such ways.

1. Let $n = m^2$; that is, dividing the circle successively into 3, 4, 5, . . . parts, we divide the altitude into 9, 16, 25, . . . parts, respectively. The area of the polyhedral surface $S_m$ (now it depends only on the index $m$) will be expressed by the formula

$$S_m = 2mR\sin\frac{\pi}{m}\sqrt{H^2 + 4m^4R^2\sin^4\frac{\pi}{2m}}.$$

We now take the limit as $m \to \infty$ (whereby $\frac{\pi}{m}$ and also $\sin\frac{\pi}{m}$ and $\sin\frac{\pi}{2m}$ tend to zero). Wishing to apply the last of the formulas given

in the footnote on page 39, we modify the expression for $S_m$ as follows:

$$S_m = 2\pi R \frac{\sin \frac{\pi}{m}}{\frac{\pi}{m}} \sqrt{H^2 + \frac{1}{4}\pi^4 R^2 \left( \frac{\sin \frac{\pi}{2m}}{\frac{\pi}{2m}} \right)^4},$$

after which we find that

$$\lim_{m \to \infty} S_m = 2\pi R \sqrt{H^2 + \frac{1}{4}\pi^4 R^2}.$$

This limit is evidently greater than $2\pi RH$, the generally known formula for the area of the lateral surface of a cylinder. We can obtain other values for the limit, as great as desired; for instance, putting $n = km^2$, where $k$ is a positive integer, we have

$$\lim_{m \to \infty} S_m = 2\pi R \sqrt{H^2 + \frac{k^2}{4}\pi^4 R^2}.$$

2. Let $n = m^3$, so that as compared with the preceding method, the number of divisions along the altitude increases still more rapidly. In the formula for $S_m$ this will merely cause the appearance of the additional factor $m^2$ in the second term of the sum under the radical sign. As a result, this term, and with it $S_m$, will tend to infinity as $m \to \infty$. This means that it is possible to establish a law for inscribing the polyhedral surfaces such that their area increases without bound and does not tend to any limit. It would appear that the area of the lateral surface of a cylinder is not a well-defined quantity.

We have arrived at an obvious absurdity and must now attempt to find the mistake.

EXAMPLE 15. *The area of a sphere of radius R is equal to* $\pi^2 R^2$.

Let us examine the hemisphere (Fig. 33) whose center lies at $O$, whose "equator" is $q$, and whose "pole" is $P$. This means that the radius $OP$ is perpendicular to the plane of the equator $q$ passing through $O$. Divide the equator $q$ into a very large number of equal parts, say $n$, and join $P$ to all the points of division by arcs of great circles. (Each arc is $\frac{1}{4}$ of an entire "meridian.") The hemisphere will then be divided into $n$ very narrow spherical triangles, each of

which is bounded by a small arc of the equator and two arcs of meridians. Some of these triangles are depicted in the figure; one of them, $PAB$, is shaded. By an increase in the number of divisions $n$, these spherical triangles may be made as narrow as desired, and

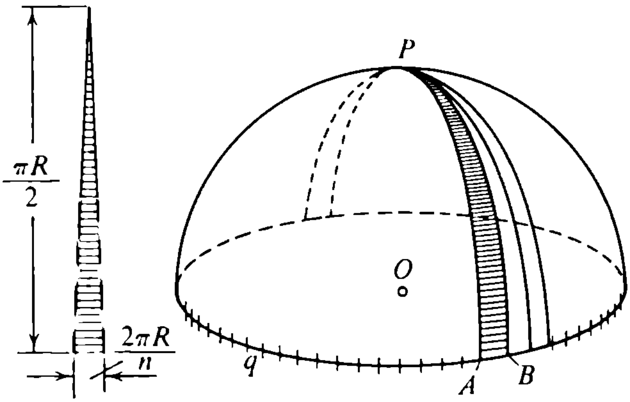



Fig. 33

the "infinitely narrow" curved triangle can be flattened out, or, as we say, "applied" to the plane with all its dimensions preserved (lengths, angles, and area). A plane triangle (isosceles) is obtained; its base is the straightened arc of length $\frac{2\pi R}{n}$, and its altitude is the straightened arc which is equal to one quarter of the circumference, that is, $\frac{\pi R}{2}$ (see the shaded triangle in Fig. 33, left). The area of such a triangle is

$$\frac{1}{2}\cdot\frac{2\pi R}{n}\cdot\frac{\pi R}{2}=\frac{1}{2n}\pi^2R^2;$$

consequently, the total area of all the $n$ triangles filling up the hemisphere is equal to $\frac{1}{2}\pi^2R^2$, and the area of the surface of the whole sphere will be $\pi^2R^2$. This contradicts the generally-known formula according to which this area is equal to $4\pi R^2$.

# 4. Analysis of the Examples Given in Chapter 3

EXAMPLE 11. In its classical form this sophism belongs properly not to geometry but to mechanics—more exactly to kinematics, the study of motion—inasmuch as the problem deals with a wheel moving in a certain manner. On the other hand, we can foresee that under close examination the kinematic label will prove to be purely superficial, since time plays no essential part (it is immaterial, for instance, whether the wheel rolls rapidly or slowly). The entire sophism can be stated in geometric language, as will be done later on.

Undoubtedly the weak side of our reasoning lies in the vagueness of the expression "the circle rolls without slipping along a straight line." It is only necessary to agree upon the precise meaning of this phrase in order to discover immediately that if one of the circles which are rigidly attached to each other rolls in this sense, then the other does not roll without slipping, and the sophistic proof collapses.

We shall first speak in the language of kinematics. That the circle rolls along the straight line without slipping means that the circle moves in such a way that it is in contact with the straight line at any given moment. The point of the circle at which the contact occurs thus has velocity zero at the given moment. In other words, the point of contact serves as an "instantaneous center of rotation" for the rolling circle. This means that at any given moment the velocity of any arbitrary point connected with the circle (not necessarily lying on the circumference of the circle) is that which it would possess if the circle were rotating about its point of contact. In particular, the direction of this velocity is perpendicular to the straight line which connects the given point with the point of contact. Thus, retaining the notation of Fig. 24, the direction of the velocity of a point arriving at the position $M'$ lies along the perpendicular to the straight line $M'P$. Hence, the straight line $M'P$ is the perpendicular (sometimes called the *normal*) to the cycloid at the point $M'$ in Fig. 26.

If, on the contrary, that point of the circle which at a given moment is in contact with the line $MN$ has a velocity different from zero, we say that the motion takes places "with positive slippage" if this velocity is in the direction of the motion, or "with negative slippage" if it is in the opposite direction. Only if there is complete absence of any slippage is it possible to affirm that the path traveled along the straight line in a given time interval is equal to the length of the circular arc corresponding to the central angle through which an arbitrary radius of the circle has turned during this time interval. For instance, in Fig. 24 and 26, $MP = \overset{\frown}{PM'}$, and $MM^*$ is equal in length to the whole circumference of the rolling circle. In case of positive slippage, $MP > \overset{\frown}{PM'}$; in case of negative slippage, $MP < \overset{\frown}{PM'}$.

We are now in a position to describe the different aspects of rolling in purely geometric terms, although for the sake of clarity we shall occasionally revert to the language of kinematics. Consider the line segment $MM^* = 2\pi R$ (Fig. 24 and 26), and at every point $P$ of it construct a tangent circle with center $O'$ lying on a given side of $MM^*$ and having a radius $R$. On each circle lay off the arc $\overset{\frown}{PM'}$ equal in length to the line segment $PM$ and on the same side of $P$ as the line segment $PM$.[1] If we carry out this construction for all possible positions of the point $P$ on the line segment $MM^*$, we shall say (returning to the language of kinematics, but remaining essentially in the realm of geometry) that the set of all the tangent circles is obtained as *the result of one revolution of the circle of radius R rolling without slipping along the straight line MN*, and the locus of all the points $M'$ corresponding to different positions of the point $P$, will be called the *trajectory* of the point $M$. If in the preceding construction we replace the equality $MP = \overset{\frown}{PM'}$ by the proportionality $MP = k\overset{\frown}{PM'}$ (where $k$ is a constant factor different from 1), we say that the circle rolls "with constant slippage coefficient $k$"; the slippage will be positive or negative depending on whether $k$ is greater or less than 1.

Armed with these precise definitions, we now turn to Aristotle's wheel. From the kinematic point of view, we know that if the larger of the concentric circles depicted in Fig. 24 rolls along the

[1] This means that the arc $PM'$ and the line segment $PM$ lie on the same side of the diameter $PO'$. In the case of a rolling curve different from a circle we would say "on the same side of the normal."

straight line $MN$ without slipping, the smaller one will not roll along the straight line $M_1N_1$ in the same manner. Actually, if the smaller circle were to roll without slipping, then at the moment when the common center of the circles is at the point $O'$, the moving figure would simultaneously have two instantaneous centers of rotation $P$ and $P_1$, and the velocity of the point $M'$ would be in a direction perpendicular to both $PM'$ and $P_1M'$, which is impossible. Apart from this, the smaller circle rolls with positive slippage, since we always have $M_1P_1 = MP = \overset{\frown}{PM'}$, and, therefore, $M_1P_1 > \overset{\frown}{P_1M_1'}$. Conversely, if the smaller circle rolls along $M_1N_1$ without slipping, the larger circle carried along by it would roll with negative slippage.

We arrive at the same result if we start from a geometric definition. If the larger circle in Fig. 24 "rolls" so that in any arbitrary position $MP = \overset{\frown}{PM'}$, then $M_1P_1 = \frac{R}{r}\,\overset{\frown}{P_1M_1'}$, where $R$ and $r$ are the radii of the larger and the smaller circle, respectively. Thus, the smaller circle rolls with positive slippage of coefficient $\frac{R}{r}\,(> 1)$. On the other hand, if the smaller circle rolls without slipping, the larger one will roll with negative slippage of coefficient $\frac{r}{R}\,(< 1)$.

*Hint following Example 11.* Let us consider two concentric and homothetic $n$-gons $A_1A_2 \ldots A_n$ and $a_1a_2 \ldots a_n$ with center $O$. (In Fig. 34, where $n = 8$, the larger polygon retains the notation of Fig.

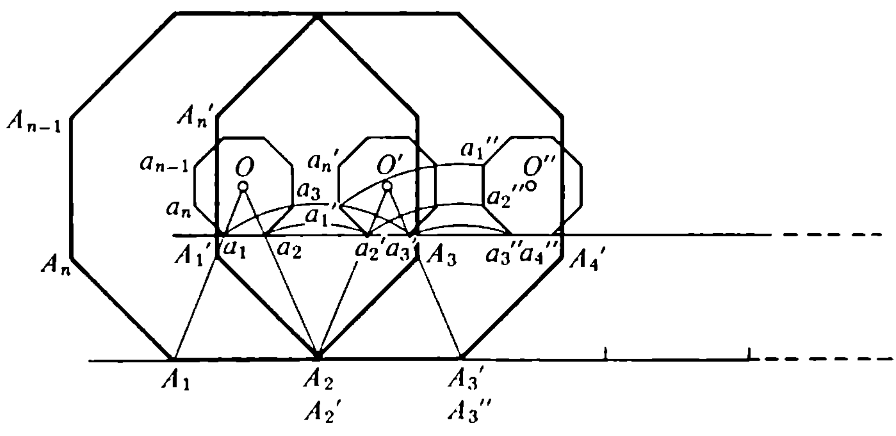


Fig. 34

25.) Let the larger polygon roll in the manner depicted in Fig. 25; at first the vertex $A_2$ remains stationary, serving as the center of

rotation until the polygon has rotated through the angle $\frac{2\pi}{n}$. (For comparison we recall that for the rolling wheel the "lowest" point also serves as the center of rotation, but only momentarily.) As a result of such a rotation the larger polygon will lie on the straight line on another of its sides, $A_2A_3$, whose new position $A_2'A_3'$ in the figure forms a direct extension of the side $A_1A_2$; owing to this the perimeter of the polygon is "unrolled" along the straight line as the rolling proceeds.

The motion undergone meanwhile by the smaller polygon, which is attached to the larger one, will be substantially different. It will also rotate about the center $A_2$ through an angle of $\frac{2\pi}{n}$, as a result of which the side $a_2a_3$ will take up the position $a_2'a_3'$. However, the position $a_2'a_3'$ is not immediately adjacent to the position $a_1a_2$. In contrast to the polygon $A_1A_2 \ldots A_n$, which is "tipped" from one side to the next, the polygon $a_1a_2 \ldots a_n$ simultaneously "tips" and "jumps" from one position to the next. In Fig. 34 two successive positions of the larger polygon and three positions of the smaller polygon are shown; the initial sections of the paths for the vertices $a_1$, $a_2$, $a_3$ are depicted. Each of these sections of the path is formed from circular arcs whose magnitude in radians is $\frac{2\pi}{n}$; for instance, the path for the vertex $a_2$ begins with $\overset{\frown}{a_2a_2'}$ with center $A_2$ and $\overset{\frown}{a_2'a_2''}$ with center $A_3'$. As in Fig. 25, we suggest that the reader not attach any significance to some of the peculiarities of Fig. 34 which are due to the particular value $n = 8$ taken there; he should make another drawing, for instance, for $n = 5$. Owing to the "jumps," the polygon $a_1a_2 \ldots a_n$, which moves along the straight line $a_1a_1^{(n-1)}$ covers a distance which is greater than its perimeter; that is an approximation of a model of "rolling with positive slippage." We suggest that the reader find out how the polygon $A_1A_2 \ldots A_n$ would move if the polygon $a_1a_2 \ldots a_n$ were made to roll along the straight line without slipping. It may be predicted that after each turn through the angle $\frac{2\pi}{n}$ about the vertex of the smaller polygon, the side of the larger one will partly overlap the preceding side of this polygon, with the result that the path it travels during a complete revolution will be less than its perimeter. We thus obtain an approximation of "rolling with negative slippage."

PROBLEM FOLLOWING EXAMPLE 11. It suffices to give a drawing and some intermediate results.

Fig. 35 depicts the path $A_1A_1'A_1'' \ldots A_1^{(n-1)}$ of the vertex $A_1$ of a rolling

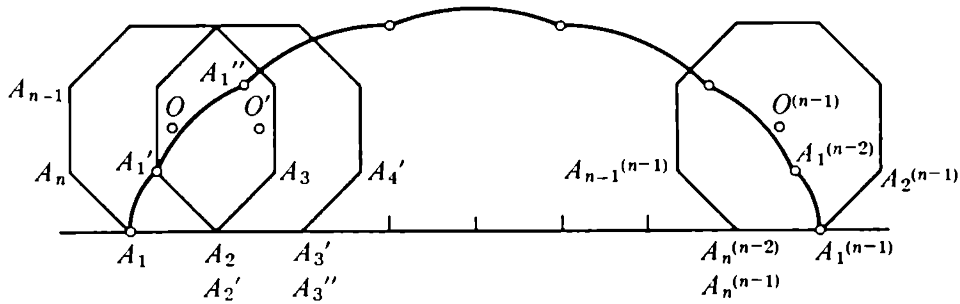



Fig. 35

polygon (in this case $n = 8$), described in greater detail on page 39. The length of this path, which consists of circular arcs, is equal to

$$\frac{2\pi}{n}\cdot 2R\left(\sin\frac{\pi}{n} + \sin\frac{2\pi}{n} + \cdots + \sin\frac{(n-1)\pi}{n}\right) = \frac{4\pi R}{n}\,\frac{\cos\dfrac{\pi}{2n}}{\sin\dfrac{\pi}{2n}}.$$

The area of the figure bounded by this trajectory and by the line segment $A_1A_1^{(n-1)}$ consists of (1) the areas of the circular sectors $A_2A_1A_1'$, $A_3'A_1'A_1''$, $\ldots$, $A_n^{(n-2)}\,A_1^{(n-2)}\,A_1^{(n-1)}$ with central angles of $\frac{2\pi}{n}$; the sum of these areas is equal to $\frac{4\pi R^2}{n}\left(\sin^2\frac{\pi}{n} + \sin^2\frac{2\pi}{n} + \cdots + \sin^2\frac{(n-1)\pi}{n}\right) = 2\pi R^2$; and (2) the areas of the triangles $A_1'A_2'A_3'$, $A_1''A_3'A_4''$, $\ldots$, $A_1^{(n-2)}\,A_{n-1}^{(n-3)}\,A_n^{(n-2)}$, whose sum is

$$2R^2\sin\frac{\pi}{n}\left(\sin\frac{2\pi}{n}\sin\frac{\pi}{n} + \sin\frac{3\pi}{n}\sin\frac{2\pi}{n} + \cdots + \sin\frac{(n-1)\pi}{n}\sin\frac{(n-2)\pi}{n}\right) = nR^2\sin\frac{\pi}{n}\cos\frac{\pi}{n}.$$

EXAMPLE 12. The logical mistake lies in the last part of the proof (preceding note 1); it is the use of the word "limit" in two entirely different senses. In one case a sequence of lines is being considered—here it is the "saws" with a variable number of "teeth"—whose points approach infinitely close to some given line. In the other case we are talking about a sequence of numbers—the lengths of the saws—which approach infinitely close to some given number. (As for the question of whether it is possible to regard the lengths of the saws as forming a sequence, see Example 12, note 2.)

*The fact that a sequence of lines tends (in the first sense) to a given line gives us no basis for the conclusion that the sequence of the lengths of the lines tends (in the second sense) to the length of the given line.*

We must not be confused by the fact that for very fine teeth the saw becomes practically indistinguishable from the straight line segment; this is not a geometric fact but a physical or even a physiological one, depending on the characteristics of our eyesight. (A powerful microscope would alter the situation.) But if we reinforce what we see by reason, the matter will present itself in the following form. It is true that in each small triangle ("sawtooth") the difference between the sum of the legs and the hypotenuse is insignificant, but the number of such triangles is very great and obviously even the smallest of terms may, when taken in very large numbers, give an appreciable sum. If we want to penetrate more deeply into the substance of the matter, we turn our attention to the fact that the sections of the saw approach the straight line $AB$ in distance but by no means in direction; however small these sections may be, they are always alternately horizontal and vertical, whereas the hypotenuse $AB$ is oblique (Fig. 27).

EXAMPLE 13. The mistake is of the same type as that in the preceding example. The sequence of wavy lines approaches infinitely close to the straight line segment, but the limit of their lengths is not the length of this segment. As before, one line (wavy in our case) approaches another in distance, but not in direction; if the line segment $AB$ is horizontal, the direction of the wavy line will always oscillate between the horizontal and the vertical, however small its circular arcs.

EXAMPLE 14. Although the picture is considerably more complex than in the two preceding examples, the logical nature of the sophism is the same: the polyhedral surface does in fact approach infinitely close to the cylindrical surface, but from this it does not follow in any way that the area of the polyhedral surface approaches infinitely close to the area of the cylindrical surface. In order to obtain a clearer insight into the connection between the two approximations, we note that it is possible, even by the method of inscribing a polyhedral surface shown in Fig. 31, to obtain the correct formula for the area of the lateral surface of the cylinder. For

instance, we can put $n = m$ or $n = 10m$; in general we can take the number of divisions along the altitude and along the circle as proportional to each other. For instance, for $n = 10m$ we have

$$S_m = 2mR \sin\frac{\pi}{m}\sqrt{H^2 + 400m^2R^2\sin^4\frac{\pi}{2m}}$$

$$= 2\pi R\frac{\sin\frac{\pi}{m}}{\frac{\pi}{m}}\sqrt{H^2 + \frac{25\pi^4R^2}{m^2}\left(\frac{\sin\frac{\pi}{2m}}{\frac{\pi}{2m}}\right)^4},$$

and for $m \to \infty$ we obtain $\lim S_m = 2\pi RH = S$.

How are we to explain the fact that if we use a different law relating $m$ and $n$, say, $n = m^2$, the area $S_m$ of the polyhedral surface tends to a limit which is greater than $2\pi RH$, while for $n = m^3$ it even tends to infinity? Let us take the liberty of answering not in the language of mathematics, foregoing any pretensions of giving a proof, but singling out only what lies within the scope of visual ideas (which may, after being analyzed mathematically, become the basis of a proof). If $n = m$, or $n = 10m$, etc., the density of the division points on the circle and on the altitude will increase at the same rate. As a result, the polyhedral surface, although it is not convex, has almost vertical faces if the cylinder is placed vertically. (We suggest that the reader prove by means of Fig. 32 that the face $MNP$ will form an angle with the horizontal plane $MKN$ which tends to $\frac{\pi}{2}$ as $m \to \infty$.) Thus, the polyhedral surface approaches the cylindrical one not only in distance, but also in direction. A different picture is obtained when $n = m^2$, or $n = m^3$, etc.; now the division points along the altitude increase in density much faster than those on the circle. As a result, the polyhedral surface becomes significantly more "indented," and in consequence a surplus area is obtained. The triangular faces will now no longer tend to become vertical; it can be shown that for $n = m^2$, the angle $PLK$ (Fig. 32) approaches infinitely close to some acute angle, and for $n = m^3$ it even tends to zero as $m \to \infty$; that is, the faces tend to become horizontal.

In conclusion, we shall examine one problem which arises naturally: Why is there no analogy between inscribing broken lines in a curved line and inscribing polyhedral surfaces in a curved surface? Why is it that in the first

case the concentration of the vertices guarantees that the lines approximate the arcs they span not only in distance but also in direction, while in the second case there may not be an approximation in direction? Without entering into details, we note only the following facts. When one point on a curve approaches a fixed point on the curve, the limit of the straight line which connects these points will be the tangent to the curve at the fixed point. When two points on a curved surface approach a third fixed point on the surface (assume the three points are never collinear), then the plane which is determined by these points does not necessarily approach the tangent plane. To verify this it is sufficient to imagine an arbitrary circle drawn on a sphere and to take on it one fixed point and two others which approach infinitely close to the first; the plane of the three points will always be the plane of the circle.

EXAMPLE 15. Before us we have an abuse of expressions which do not have mathematical significance, such as "very large number," "very narrow triangle," "small arc," "infinitely narrow triangle"; these expressions are appropriate when attempting to give a visual description of a geometric figure—this method of description has been applied in several examples above—but they are completely unsuitable as tools for giving a proof or deriving a formula. The exact mistake lies in the assertion that an infinitely narrow triangle may be applied to a plane, that is, that it is possible to replace it by a plane triangle whose sides, angles, and areas have the same magnitude as those of the spherical triangle. In fact, it is not possible to apply any spherical triangle, however small, onto a plane in this sense. This is evident from the fact that the sum of the angles of a plane triangle is always equal to 180°, whereas for a spherical triangle it is always greater than 180°. In our example the angles $A$ and $B$ of the spherical triangle $PAB$ are right angles (Fig. 33, right); if such a triangle could be applied to a plane, we would obtain a plane isosceles triangle (Fig. 33, left) with two right angles at the base.